

Exercices de Statistique  
HEI 3 - 2014/2015

A. RIDARD



# Table des matières

1	Loi normale	5
2	Estimation	7
3	Tests statistiques	11
4	Régression linéaire	13
5	Analyse de variance	19
	Annales	25



# Chapitre 1

## Loi normale

### Exercice 1.

La demande mensuelle d'un produit obéit à une loi normale. Elle a une probabilité 0,1 d'être inférieure à 15 000 unités (resp. supérieure à 25 000 unités).

1. Quels sont les paramètres de cette loi ?
2. Calculer la probabilité qu'en un mois la demande dépasse 23 000 unités.
3. Quel doit être le stock pour ne risquer une rupture qu'avec une probabilité d'environ 0.1% ?

### Exercice 2.

Une société envisage la mise en place de nouveaux équipements. Dans le cadre de ce projet, elle a défini trois tâches A, B et C. On sait que la tâche A dure 10 semaines et que les tâches B et C ont des durées aléatoires indépendantes ; la durée de B obéissant à la loi normale de moyenne 18 et d'écart-type 4, la durée de C obéissant à la loi normale de moyenne 20 et d'écart-type 5.

1. Quelle est la probabilité que la tâche B (resp. C) dure entre 18 et 25 semaines ?
2. Quelle est la probabilité que la durée totale D des trois tâches ne dépasse pas 55 semaines ?
3. Déterminer un intervalle centré sur l'espérance mathématique dans lequel se trouve la durée totale avec une probabilité de 0,75.

### Exercice 3.

Les ampoules de la marque A ont une durée de vie moyenne de 2500 heures avec un écart-type de 500 heures, celles de la marque B ont une durée de vie moyenne de 2300 heures avec un écart-type de 800 heures. On prélève 300 ampoules A et 200 ampoules B.

1. Quelle est la loi de la durée de vie moyenne des 300 ampoules A (resp. des 200 ampoules B) ?
2. Quelle est la probabilité que la durée de vie moyenne des 300 ampoules A ne soit pas supérieure de plus de 100 heures à la durée de vie moyenne des 200 ampoules B ?
3. Quelle est la probabilité que l'écart entre les deux durées de vie moyenne ne dépasse pas 40 heures ?

### Exercice 4 (juin 2009).

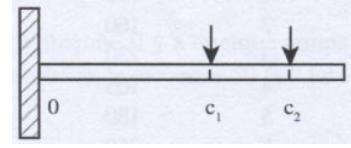
Un balladeur mp3 fabriqué par la compagnie Multisonic est garanti contre tout défaut de fabrication pour une période de 2 ans. D'après l'expérience de la compagnie, les chances d'observer une non-conformité majeure durant les 26 mois (respectivement 52 mois) suivant l'achat sont de 1 sur 100 (respectivement 975 sur 1000). Supposons que le temps  $X$  requis après l'achat pour qu'une non-conformité majeure se présente soit distribué normalement.

1. Déterminer les paramètres de cette gaussienne.
2. Quelle est la probabilité que l'appareil présente une non-conformité majeure avant la fin de la période de garantie ?

3. Quelle devrait être la période de garantie si Multisonic ne souhaitait remplacer que 0,05% des appareils vendus ?

**Exercice 5** (juin 2009).

Si deux charges sont appliquées à une poutre en porte-à-faux selon le schéma ci-contre, le moment fléchissant à O dû aux charges est  $M_F = c_1 X_1 + c_2 X_2$ .



Supposons que  $c_1 = 2$  mm, que  $c_2 = 6$  mm et que  $X_1, X_2$  soient des variables aléatoires indépendantes respectivement de moyenne 5 kN et 8kN, d'écart-type 0,25 kN et 0,40 kN.

1. Déterminer l'espérance et l'écart-type du moment fléchissant. Soit  $X, Y$  des variables aléatoires indépendantes et  $a, b$  des réels. Alors

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \text{ et } \text{Var}(aX + b) = a^2 \text{Var}(X)$$

2. Si  $X_1$  et  $X_2$  sont distribuées normalement, quelle est la probabilité que le moment fléchissant soit supérieur à 64 Kn.mm ?

**Exercice 6** (janvier 2010).

La vitesse (km/h) des voitures passant à un certain point d'une route peut être considérée comme une variable aléatoire de loi normale. Par observation, on trouve que 95% des voitures a une vitesse inférieure à 150 km/h et 10% a une vitesse inférieure à 90 km/h.

1. Déterminer les paramètres de cette loi.
2. Calculer le pourcentage de voitures roulant entre 120 et 140 km/h.

**Exercice 7** (juin 2010).

Un bar débite la bière en chopes dont le contenu effectif est une variable aléatoire  $X$  supposée gaussienne de moyenne  $m = 25$  cl et d'écart-type  $\sigma = 2$  cl.

1. Déterminer la probabilité que votre chope de bière contienne :
  - (a) plus de 26 cl de bière.
  - (b) moins de 23 cl de bière.
2. Déterminer la probabilité pour qu'il y ait moins de 1 cl d'écart entre votre chope et celle de votre ami.

**Exercice 8** (janvier 2011).

Une machine remplit automatiquement des boîtes de sucre en poudre de telle façon que le poids de sucre effectivement contenu dans une boîte soit une variable aléatoire normale de paramètres  $m$  et  $\sigma$  exprimés en grammes. On souhaite régler la machine de sorte que le poids de sucre contenu dans une boîte dépasse 980 grammes avec une probabilité de 95%.

1. Lorsque  $\sigma = 30$ , quelle valeur faut-il donner à la moyenne  $m$  ?
2. Lorsque  $m = 1000$ , quelle valeur doit avoir l'écart-type  $\sigma$  ?

**Exercice 9** (janvier 2011).

Une entreprise décide de construire des sièges de bureau réglables en hauteur. Si  $h$  est la hauteur du siège par rapport au sol dans sa position la plus basse, alors  $2h$  est la hauteur du siège dans sa position la plus élevée.

Afin de déterminer la valeur optimale de  $h$ , l'entreprise a effectué une enquête auprès de sa clientèle et a mesuré pour chaque individu, la hauteur  $H$  du siège qui lui paraissait la plus confortable.

L'enquête a montré que  $H$  pouvait être ajustée par une loi gaussienne, autour d'une espérance  $m = 48$  cm.

1. Quelle est la hauteur  $h$  que l'entreprise doit choisir de façon à satisfaire le maximum de sa clientèle ?
2. Quelle serait alors la proportion de clientèle satisfaite si  $\sigma = 8$  ?

## Chapitre 2

# Estimation

### Exercice 10.

Un fabricant et livreur de pizzas à domicile désire faire une étude sur le temps moyen qui s'écoule entre le moment où la commande est passée par téléphone et le moment où le client est livré. Une observation rapide faite sur 25 commandes fait ressortir un temps moyen de 27 mn.

En supposant que le temps de fabrication-livraison suive une loi normale d'espérance  $m$  et d'écart-type 1, donner un intervalle de confiance pour  $m$  au seuil 95%.

### Exercice 11.

On suppose que le chiffre d'affaires journalier  $X$  d'un hypermarché est distribué selon une loi normale d'espérance  $m$  et d'écart-type inconnu. Durant 31 jours, on a relevé la valeur du chiffre d'affaires (en milliers d'euros) et obtenu le résultat suivant :

$$\sum_{i=1}^{31} x_i = 279 \text{ et } \sum_{i=1}^{31} x_i^2 = 2945$$

Donner un intervalle de confiance pour  $m$  au seuil 95%.

### Exercice 12.

On vous demande de faire une enquête afin de déterminer le pourcentage des votants qui exerceront leur droit de vote lors du prochain référendum. On exige de vous une estimation du pourcentage au seuil 95% avec une incertitude de  $\pm 2\%$ .

Quel est le nombre minimal de personnes que vous devez interroger ?

### Exercice 13.

Une étude réalisée sur la vitesse coronarienne  $X$  fournit, pour 18 individus, les résultats suivants :

$$75, 77, 78, 77, 77, 72, 72, 72, 70, 71, 69, 69, 68, 66, 64, 66, 62, 61$$

On en tire :  $\bar{x} = 70.33$  et  $s = 5.09$

1. Donner une estimation ponctuelle de la moyenne et de la variance de  $X$ .
2. On suppose que  $X \sim \mathcal{N}(m; \sqrt{26})$ .  
Déterminer l'intervalle de confiance de  $m$  au niveau 0,98.
3. On suppose maintenant que  $X \sim \mathcal{N}(m; \sigma)$ .
  - (a) Déterminer l'intervalle de confiance de  $m$  au niveau 0,98.
  - (b) Déterminer l'intervalle de confiance de  $\sigma^2$  au niveau 0,98.

**Exercice 14** (juin 2009).

L'entreprise Metallo a effectué des essais en traction sur des tiges métalliques employées dans l'assemblage de structures. Notons  $X$  la résistance à la traction (en kg/cm<sup>2</sup>) d'espérance  $m$  et de variance  $\sigma^2 = 900$ . Des essais sur 50 tiges fournissent un échantillon de moyenne 380 kg/cm<sup>2</sup>.

- Notons  $X_i$  la résistance à la traction de la  $i$ -ième tige de l'échantillon et  $\bar{X} = \frac{1}{50} \sum_{i=1}^{50} X_i$ . Déterminer une approximation de la loi de  $\frac{\sqrt{50}}{\sigma}(\bar{X} - m)$ .
- En déduire un intervalle de confiance pour  $m$  au seuil 95%.

**Exercice 15** (juin 2009).

Un relevé effectué par le responsable du service informatique d'une entreprise multinationale indique l'utilisation suivante (en minutes par heure) pour neuf terminaux choisis au hasard :

28,3	26,4	27,0	22,5	23,5	29,1	26,8	26,7	30,9
------	------	------	------	------	------	------	------	------

On en tire :  $\bar{x} = 26.8$  et  $s^* = 2.6$

- Quelle hypothèse devons-nous faire pour déterminer un intervalle de confiance au seuil 95% pour la moyenne d'utilisation de tous les terminaux de cette entreprise ? Préciser alors cet intervalle.
- En conservant cette hypothèse, déterminer un intervalle de confiance au seuil 95% pour la variance.

**Exercice 16** (juin 2010).

A partir des vitesses relevées ci-dessous, estimer par intervalle de confiance au niveau 95%, dans les deux cas suivants, la vitesse moyenne de tous les véhicules circulant en France en supposant la vitesse gaussienne de moyenne  $m$  et d'écart-type  $\sigma$ .

Vitesse de 20 véhicules en km/h									
124	105	94	93	106	101	90	91	118	115
100	96	112	109	89	102	95	89	88	108

On en tire :  $\bar{x} = 101.25$  et  $s^* = 10.62$

- Si  $\sigma = 9$ .
- Si  $\sigma$  est inconnu.

**Exercice 17** (juin 2010).

Les dernières campagnes de contrôle, effectuées chez un grand brasseur, laissent supposer que la précision des remplisseuses les plus anciennes n'est plus égale à sa valeur initiale de 0,1 cl.

A partir des contenances relevées ci-dessous, estimer par intervalle de confiance au niveau 95 %, l'écart-type de la production totale en supposant la contenance des bouteilles gaussienne.

Contenance de 20 bouteilles en cl									
74,8	75	74,9	75,2	75	74,9	74,9	75	75	75,1
75	74,8	74,9	74,8	75	75	75,1	75	75,1	75,1

On en tire :  $s^2 = 0.0116$

**Exercice 18** (janvier 2011).

En mai 2006, le ministre de l'Economie des Finances et de l'Industrie a mis en place la déclaration de revenus préremplie par l'administration. Les premiers contrôles effectués sur un échantillon de 780 déclarations d'imposition préremplies montrent que 12% des ces déclarations présentent au moins une erreur. Le montant moyen des erreurs, calculé à l'euro près, étant de 3500 euros. On supposera l'écart-type  $\sigma$  connu égal à 800 euros.

1. Donner un intervalle de confiance au niveau 95% pour les paramètres suivants :
  - (a) La proportion  $p$  de déclarations erronées.
  - (b) Le montant moyen  $m$  des erreurs.
2. Donner le nombre minimal de déclarations à contrôler pour avoir au niveau 95% :
  - (a) une incertitude n'excédant pas 2% pour la proportion  $p$  de déclarations erronées.
  - (b) une incertitude n'excédant pas 50 pour le montant moyen  $m$  des erreurs.

**Exercice 19** (janvier 2011).

Voici les détente sèches (en cm) de 30 basketteurs représentatifs des basketteurs du Nord. La détente sèche sera supposée gaussienne de moyenne  $m$  et d'écart-type  $\sigma$ .

Détente	57	58	59	60	61	62	63	64	65
Effectif	1	5	1	5	3	7	2	2	4

On en tire :  $\bar{x} = 61.23$  et  $s^2 = 5.52$

1. Estimer par intervalle de confiance au niveau 98% les paramètres suivants (des basketteurs du Nord) :
  - (a) L'écart-type de la détente sèche.
  - (b) La détente sèche moyenne.
2. En supposant maintenant  $\sigma$  connu égal à 2.37, estimer par intervalle de confiance au niveau 98% la détente sèche moyenne des basketteurs du Nord.
3. Comparer les intervalles de confiance pour  $m$  et commenter.
4. Refaire les questions 1.(b) et 2. sans supposer la détente sèche gaussienne.



# Chapitre 3

## Tests statistiques

### Exercice 20.

Le cahier des charges du principal client d'un fabricant de cartes à puces impose une proportion de cartes défectueuses dans la production totale, inférieure ou égale à 0,5%. Suite à une livraison contenant 0,7% de cartes défectueuses, selon le client, le fournisseur décide de contrôler un échantillon de taille 1000.

1. Six cartes étant défectueuses, tester  $\begin{cases} H_0 : p = 0,005 \\ H_1 : p = 0,007 \end{cases}$  au seuil de 2% ?
2. Calculer le risque de deuxième espèce puis commenter.

### Exercice 21.

Deux échantillons indépendants et gaussiens ont donné les résultats suivants :

$$n_A = 40, \bar{x}_A = 25.2, s_A^* = 5.2 \text{ et } n_B = 50, \bar{x}_B = 22.8, s_B^* = 6$$

La différence observée entre les deux moyennes empiriques est-elle statistiquement significative au seuil de 5% ?

### Exercice 22 (juin 2010).

Les associations de consommateurs font appel à des organismes indépendants pour tester les caractéristiques avancées par les fabricants pour de nombreux produits. Une des questions soulevées est relative à l'influence, sur la consommation d'électricité, de l'utilisation d'un adoucisseur d'eau pour alimenter la machine à laver. L'entartrage n'intervenant qu'après une utilisation prolongée, les 42 machines testées sont âgées de 4 ans.

A partir des relevés ci-dessous, décider si l'efficacité de l'adoucisseur est statistiquement significative au seuil de 5% en supposant les consommations avec adoucisseur ( $X_A$ ) et sans adoucisseur ( $X_B$ ) gaussiennes et indépendantes.

*Indice : on effectuera pour cela un test unilatéral de comparaison de moyennes précédé d'un test de comparaison de variances, les deux au seuil de 5%*

25 consommations avec adoucisseur (A) en kw/h									
1,07	0,79	0,66	0,59	0,83	0,8	0,87	0,93	0,75	0,78
0,68	0,71	0,82	0,76	0,93	0,82	0,74	0,77	0,81	0,89
0,91	1,11	0,78	0,79	0,79					
17 consommations sans adoucisseur (B) en kw/h									
1,01	0,91	0,81	0,93	0,89	0,9	0,88	1,07	1,04	0,77
0,92	0,82	0,67	0,9	0,81	0,93	0,88			

On en tire :  $\bar{x}_A = 0.815, s_A^{*2} = 0.013, \bar{x}_B = 0.891$  et  $s_B^{*2} = 0.010$



## Chapitre 4

# Régression linéaire

### Exercice 23.

Le tableau suivant donne la consommation  $X$  de graisse ainsi que le taux  $Y$  de mortalité par athérosclérose pour 100 000 habitants de la Norvège sur une période couvrant la seconde guerre mondiale.

Date	Consommation de graisse en kg par an et par personne	Taux de mortalité par athérosclérose pour 100 000 habitants
1938	14.4	29.1
1939	16.0	29.7
1940	11.6	29.2
1941	11.0	26.0
1942	10.0	24.0
1943	9.6	23.1
1944	9.2	23.0
1945	10.4	23.1
1946	11.4	25.2
1947	12.5	26.1

1. Construire le nuage de points représentant ces données.
2. Déterminer l'équation de la droite d'ajustement de  $Y$  en  $X$ .
3. Calculer le coefficient de détermination  $r^2$ .  
La consommation de graisse explique t'elle la mortalité par athérosclérose ?

### Exercice 24.

Un physicien a mesuré le volume  $V$  et la pression  $P$  d'un gaz. Ses résultats sont consignés dans le tableau suivant.

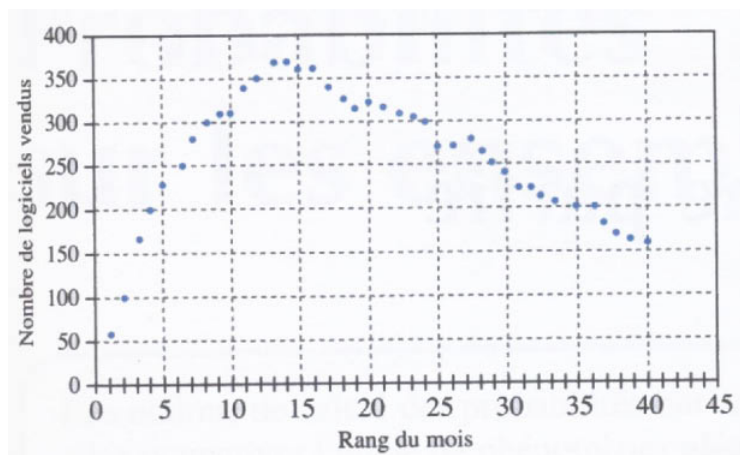
Volume en $\text{cm}^3$	1069	1216	1425	1746	2334	3818
Pression en $\text{g}/\text{cm}^3$	3778	3056	2321	1753	1185	623

Montrer qu'il existe entre  $P$  et  $V$  une relation de la forme :  $PV^\gamma = C$ .

### Exercice 25 (mars 2009).

Une entreprise réalise une étude de marché avant de commercialiser des logiciels à usages professionnels.

1. *Exploitation statistique d'un modèle passé.*  
Des concurrents ont récemment vendu un produit similaire. Le nombre de logiciels vendus chaque mois est donné par le graphique ci-contre :



Un extrait est fourni dans le tableau suivant :

Rang du mois $x_i$	1	6	11	16	21	26	31	36
Nombre de logiciels vendus $z_i$	60	250	340	360	320	270	220	200

- (a) Sans calcul, justifier qu'un ajustement linéaire n'est pas approprié.  
 (b) Reproduire et compléter le tableau suivant (les  $y_i$  seront arrondies au centième) :

Rang du mois $x_i$	1	6	11	16	21	26	31	36
$y_i = \ln\left(\frac{z_i}{x_i}\right)$	4,09						1,96	1,71

- (c) Déterminer le coefficient de corrélation linéaire  $r$  entre  $x$  et  $y$  au millième près.  
 Que peut-on en déduire ?  
 (d) Donner une équation de la droite de régression de  $y$  en  $x$  sous la forme  $y = ax + b$  où  $a$  et  $b$  sont deux réels arrondis au millième.  
 (e) En prenant des arrondis plus larges pour  $a$  et  $b$ , on obtient :

$$\ln\left(\frac{z}{x}\right) = -0,07x + 4$$

Exprimer alors  $z$  en fonction de  $x$ .

## 2. Perspectives

L'équipe commerciale envisage de mener une campagne plus dynamique, pour son nouveau produit plus complet. Le nombre mensuel des ventes serait modélisé par la fonction  $f$  définie sur l'intervalle  $[0; 36]$  par  $f(x) = 100xe^{-0,1x}$ .

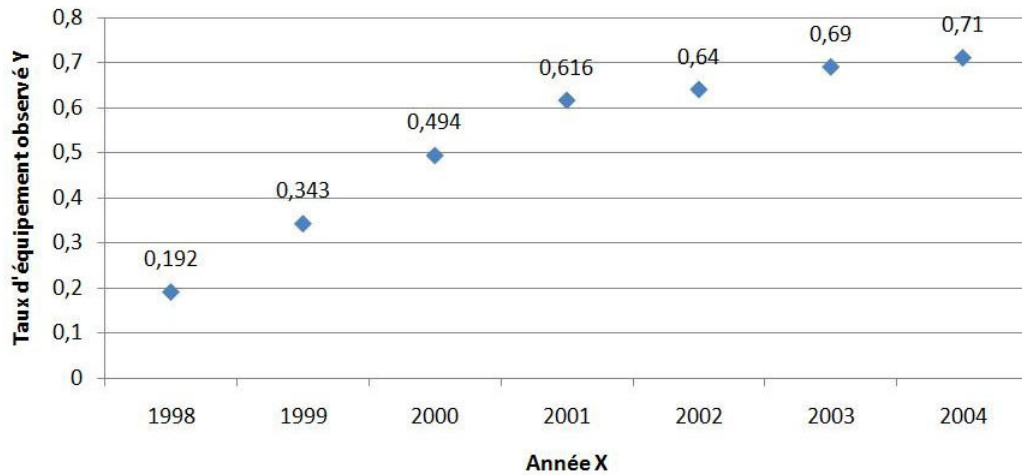
- (a) Dresser le tableau de variations de  $f$ .  
 (b) Tracer la courbe représentative de  $f$ .  
 (c) L'entreprise arrêtera la commercialisation du produit dès que le nombre de ventes repassera au-dessous de 150 unités par mois. Déterminer, à l'aide du graphique, à partir de quel mois cessera cette commercialisation.

### Exercice 26 (novembre 2010).

Les opérateurs de téléphonie mobile ont besoin de prévoir les taux de pénétration sur le marché du téléphone mobile pour les années à venir. En effet, ce sont les prévisions du marché qui leur permettent de définir les stratégies d'investissement et d'innovation.

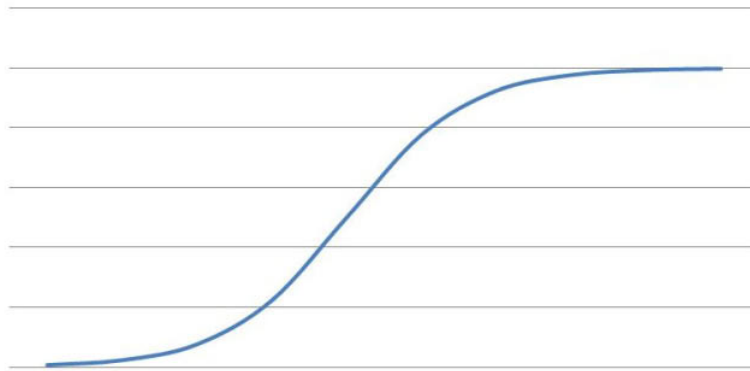
Les taux d'équipement observés depuis 1998 jusqu'à septembre 2004 sont indiqués sur le graphique suivant. En 2003, le taux de pénétration en France est de 69%, ce qui positionne la France au niveau du Japon et loin derrière l'Italie.

## Taux d'équipement en téléphones portables de 1998 à 2004 en France



Il s'agit d'ajuster le nuage de points par une courbe, afin d'utiliser celle-ci pour prévoir les taux d'équipement futurs. L'observation suggère un ajustement logistitique  $Y = \frac{k}{1 + ke^{-(aX+b)}}$ . Les fonctions logistiques sont bien adaptées pour modéliser l'évolution dans le temps des ventes d'un produit depuis sa création. En effet, il n'est pas rare que la phase de démarrage soit suivie d'une phase de croissance forte, elle même suivie d'une dernière phase de stagnation :

### Un exemple de fonction logistique



1. Montrer que l'ajustement logistique  $Y = \frac{k}{1 + ke^{-(aX+b)}}$  se ramène à l'ajustement linéaire  $Y' = aX' + b$  avec  $X' = X$  et  $Y' = \ln(\frac{kY}{k-Y})$ .
2. Déterminer  $\lim_{X \rightarrow +\infty} \frac{k}{1 + ke^{-(aX+b)}}$  quand  $a > 0$ .
3. Pour les deux valeurs suivantes de  $k$  qui représente le seuil de saturation, réaliser l'ajustement logistique en précisant bien les valeurs de  $a$  et  $b$  ainsi que le coefficient de détermination  $r^2$  :
  - (a)  $k = 0,83$
  - (b)  $k = 0,78$
4. Quel modèle doit-on privilégier pour une prévision à court terme ?

**Exercice 27** (avril 2012).

MOULICAF envisage de racheter l'entreprise RAPENEX, entreprise concurrente confrontée à de délicats problèmes de succession. Le chiffre d'affaires des derniers exercices est le suivant (en millions de livres poldaves), après correction de l'incidence de l'inflation :

Année	t	$x_t$
1998	0	126
1999	1	115
2000	2	165
2001	3	192
2002	4	176
2003	5	237

- Modéliser cette série chronologique à l'aide du modèle linéaire  $X = aT + b$  en précisant bien les paramètres.
- Donner le coefficient de détermination  $r^2$  et interpréter le résultat.
- Notons  $x_t^* = at + b$  le chiffre d'affaires estimé par le modèle et  $e_t = x_t - x_t^*$  l'écart résiduel entre le chiffre d'affaires observé et celui estimé par le modèle.
  - Déterminer les 6 écarts résiduels, leur moyenne et leur variance  $s_{x/t}^2$ .
  - Déterminer la variance  $s_x^2$  des  $x_t$  et comparer  $s_{x/t}^2$  avec  $(1 - r^2)s_x^2$ .
- Modéliser maintenant la série chronologique à l'aide des modèles suivants en précisant bien les paramètres ainsi que le modèle linéaire équivalent utilisé :
  - Le modèle quadratique  $X = aT^2 + b$
  - Le modèle exponentiel  $X = a \times b^T$
- Choisir le meilleur des trois modèles en justifiant la réponse.
- En utilisant le modèle choisi, prévoir le chiffre d'affaires de l'exercice 2004.

**Exercice 28.**

Un agent immobilier possède huit studios à louer dans une banlieue de Paris. Il a noté dans un tableau la distance  $X$  de ces studios à la station de RER la plus proche ainsi que le montant  $Y$  du loyer mensuel.

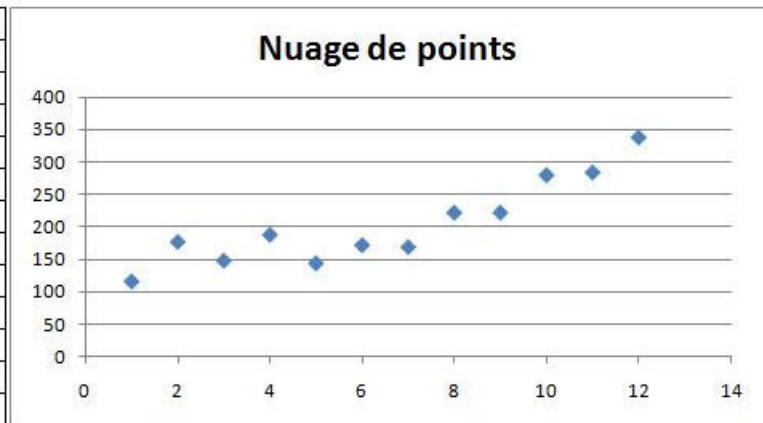
Distance en km	0,54	0,90	1,26	1,98	2,16	4,14	5,22	5,4
Loyer mensuel en Francs	3300	3000	3120	2460	2760	2280	2400	2280

- Représenter graphiquement ces données.
- Adapter à ces données un modèle linéaire puis un modèle puissance.  
Quel est le meilleur modèle ?
- Déterminer, au niveau 95%, l'intervalle de prévision du loyer pour une distance de 6 km.

**Exercice 29** (inspiré de novembre 2009).

Considérons les données trimestrielles suivantes relatives à des souscriptions de contrats d'assurance vie de fin mars 2000 à fin décembre 2002 :

t	Y= nombre de contrats souscrits
1	117
2	178
3	149
4	189
5	145
6	173
7	170
8	223
9	223
10	281
11	285
12	339



1. On ajuste cette série par le modèle linéaire :  $Y = at + b$ .  
Calculer les coefficients de ce modèle par la méthode des moindres carrés.
2. On ajuste maintenant cette série par le modèle quadratique :  $Y = at^2 + b$ .  
Calculer les coefficients de ce modèle par la méthode des moindres carrés.
3. Quel modèle choisissez-vous (justifiez votre réponse) ?
4. Déterminer, au niveau 95%, l'intervalle de prévision du nombre de contrats souscrits à la fin du mois de décembre 2003.

**Exercice 30** (inspiré de avril 2010).

Le tableau suivant présente l'évolution du taux de mortalité infantile  $M$  (pour 1000) de 1950 à 2004.

Année	T	M	Année	T	M
1950	0	51,9	1990	40	7,3
1955	5	38,6	1995	45	4,9
1960	10	27,4	2000	50	4,4
1965	15	21,9	2001	51	4,5
1970	20	18,2	2002	52	4,1
1975	25	13,8	2003	53	4
1980	30	10	2004	54	3,9
1985	35	8,3	2010		

1. Ajuster un modèle exponentiel de la forme  $M = ae^{\beta T}$  en précisant bien les paramètres.
2. La qualité de cet ajustement est-elle correcte ?
3. Déterminer, à partir de ce modèle, le taux de mortalité infantile prévu en 2010.
4. Déterminer, au niveau 95%, l'intervalle de prévision du taux de mortalité infantile en 2010.

**Exercice 31** (inspiré de avril 2011).

Dans un circuit électrique, on a relevé l'évolution temporelle de la tension aux bornes d'un appareil inconnu, repris dans le tableau ci dessous :

T (en ms)	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2.0	2.2	2.4	2.6	2.8	3.0
U (en Volt)	5.7	5.5	5.3	5.1	4.9	4.8	4.5	4.2	3.8	3.4	3.0	2.5	2	1.4	0.7

1. Représenter ce nuage de points.
2. On ajuste cette série par le modèle linéaire  $U = aT + b$ . Déterminer les coefficients de ce modèle.
3. Déterminer, au niveau 95%, l'intervalle de prévision de la tension pour un temps de 3,3 ms.

**Exercice 32** (inspiré de novembre 2011).

Le tableau ci-dessous donne l'évolution du nombre de personnes âgées en milliers ( $Y$ ) de plus de 85 ans, en France métropolitaine, de 1950 à 2000.

Année	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000
$x$	0	5	10	15	20	25	30	35	40	45	50
$y$	201	231	290	361	423	498	567	684	874	1079	1267

1. Ajuster un modèle exponentiel de la forme  $Y = \alpha e^{\beta X}$ , cet ajustement est-il correct ?
2. Déterminer, au niveau 95%, l'intervalle de prévision du nombre de personnes âgées de plus de 85 ans en 2010.

## Chapitre 5

# Analyse de variance

### Exercice 33.

Des élèves choisis au hasard dans une promotion sont répartis en trois groupes, et leurs notes à l'issue de l'épreuve de statistique sont inscrites dans le tableau suivant. Le groupe a-t-il un effet significatif sur la note ?

Groupe 1	Groupe 2	Groupe 3
19	7	11
19	7	13
17	8	20
7	11	2
13	20	1
17	16	16
7	13	5
13	6	2
8	2	8
15	19	14
3	14	6
19	12	16
7	6	2
20	10	16
14	19	6
14	20	8
9	13	11
17	2	4
19	19	13
3	20	5
8	10	20
15	3	19
6	8	10
5	4	6
19		17
9		10
19		20
6		
3		
17		
3		
6		
7		
8		
18		

**Exercice 34.**

Le tableau suivant présente les salaires annuels bruts d'individus au bout de cinq ans d'expériences selon leur niveau de formation initiale. Qu'en pensez-vous ?

Licence	Master	Doctorat
35,9	39,7	25,6
32,5	32,6	48,2
36	25,7	47,3
28,1	35,4	29,3
22,4	29,1	35,6
23,5	40,3	26,4
24,6	27,6	28,6
21,5	22,1	47,5
24,2	28,9	35,8
23,7	31,6	42,6
30,7	32,5	45

**Exercice 35.**

A partir du tableau suivant, préciser si le facteur A (resp. B) a un effet significatif.

	B1	B2	B3	B4	B5
A1	1,2	1,3	1,5	1,3	1,6
A2	1,3	1,4	1,2	1,4	1,2
A3	1,3	1,2	1,4	1,3	1,5
A4	1,2	1,1	1,3	1,5	1,4

**Exercice 36.**

- On considère le rendement de trois variétés de maïs dont la culture a été effectuée avec cinq engrais différents. Dans le tableau suivant, chaque observation est spécifique à une variété et un engrais. Il s'agit de tester le rendement moyen des parcelles en tenant compte de deux facteurs de variabilité : la variété et le type d'engrais. Effectuer les tests avec un seuil de 5%.

Engrais	Variété 1	Variété 2	Variété 3
A	256	226	200
B	205	250	170
C	199	212	180
D	156	170	190
E	193	220	200

- On dispose maintenant de deux observations pour chacun des trois échantillons de variétés de maïs et chacune des cinq catégories d'engrais. Effectuer les tests nécessaires pour décider du bien fondé d'une sélection d'une variété particulière de maïs et/ou de l'usage d'engrais. On commencera évidemment par l'étude des interactions (pourquoi ?). Le seuil des différents tests est fixé à 5%.

Engrais	Variété 1	Variété 2	Variété 3
A	183	187	205
	256	226	200
B	199	192	194
	205	250	170
C	193	196	200
	199	212	180
D	198	197	199
	156	170	190
E	187	207	190
	193	220	200

**Exercice 37.**

Un fabricant de coussinets en bronze fritté se propose de déterminer si la résistance à la rupture du bronze dépend des lots de poudre de cuivre et d'étain utilisés pour son élaboration. On réalise à partir de trois lots différents de poudre de cuivre (facteur A) et de trois lots différents de poudre d'étain (facteur B), neuf mélanges de composition identiques (90% de cuivre et 10% d'étain), correspondants aux neuf combinaisons deux à deux des lots de cuivre et d'étain utilisés. A partir de chacun de ces mélanges, on comprime, sous une même pression, quatre éprouvettes de flexion identique. Les trente-six éprouvettes obtenues sont ensuite frittées en une même opération dans un four à atmosphère réductrice. Ces éprouvettes sont enfin cassées sur une machine d'essai. Les charges de rupture ainsi déterminées arrondies à 0.12 kg /mm près sont reportées dans le tableau ci-après, en hectogrammes.

	B1	B2	B3
A1	6	1	0
	7	1	5
	3	4	6
	8	3	2
A2	1	6	0
	6	4	3
	7	4	2
	4	10	2
A3	6	8	2
	10	3	4
	8	7	3
	7	7	7

1. L'interaction de l'étain et du cuivre a-t-il un effet significatif sur la résistance au seuil de 5% ?
2. L'étain (resp. le cuivre) a-t-il un effet significatif sur la résistance au seuil de 5% ? 2% ?

**Exercice 38.**

L'influence de 4 additifs de carburant (A) sur les émissions d'oxyde d'azote doit être déterminée. De plus, il est prévu d'utiliser pour les essais, 4 véhicules (B) et 4 conducteurs (C) distincts. Les essais réalisés ont conduit aux résultats reportés dans le tableau ci-dessous :

	B1	B2	B3	B4
A1	C1 21	C2 26	C4 20	C3 25
A2	C4 23	C3 26	C1 20	C2 27
A3	C2 15	C4 13	C3 16	C1 16
A4	C3 17	C1 15	C2 20	C4 20

1. L'additif a-t-il un effet significatif sur les émissions d'oxyde d'azote au seuil de 5% ?
2. Le véhicule a-t-il un effet significatif sur les émissions d'oxyde d'azote au seuil de 5% ?
3. Le conducteur a-t-il un effet significatif sur les émissions d'oxyde d'azote au seuil de 5% ? 10% ?



# Annales



## Partiel de Probabilités et Statistique

HEI3

Novembre 2012

- Durée : 2 heures
- Autorisé : calculatrice et recto-verso manuscrit
- Le barème n'est donné qu'à titre indicatif

### Exercice 1 (2 points)

La taille moyenne des pygmées d'une tribu est de 1m40. 10% des pygmées de cette tribu font plus de 1m50. En supposant que la taille des pygmées suive une loi normale, quel est l'écart-type de cette distribution ?

### Exercice 2 (6 points)

La taille d'un épi de blé dans un champ est modélisée par une variable aléatoire  $X$  de loi normale  $N(15;6)$  (unité : le cm).

1. Quelle est la probabilité pour qu'un épi ait une taille inférieure à 16 cm ?
2. On admet qu'il y a environ 15 millions d'épis dans le champ. Donner une estimation du nombre d'épis de plus de 20 cm.
3. Quelle est la probabilité pour qu'un épi prélevé dans le champ ait une taille dans l'intervalle  $[16 ; 20]$  ?
4. On suppose que la taille d'un épi de blé d'un autre champ est modélisée par une variable aléatoire  $Y$  de loi normale  $N(10;4)$  et que  $X$  et  $Y$  sont des variables aléatoires indépendantes. Quelle est la probabilité pour qu'un épi pris dans le premier champ soit plus grand qu'un épi pris dans le second ?

### Exercice 3 (3 points)

Parmi un groupe de malades qui se plaignent de ne pas bien dormir, certains ont absorbé un somnifère sous forme de cachets, d'autres ont absorbé des cachets de sucre (*tous pensaient qu'on leur avait administré un somnifère*). Après quoi, on leur a demandé si les cachets ont été efficaces ?

Le tableau suivant indique les différentes réponses. En supposant que tous les malades aient dit la vérité, tester l'hypothèse qu'il n'y a aucune différence entre le somnifère et le sucre au seuil de signification de 0.05.

	Ont bien dormi	N'ont pas bien dormi
Ont pris le somnifère	44	10
Ont pris du sucre	81	35

### Exercice 4 (5 points)

Des études pédagogiques ont établi que pour une bonne compréhension des matières enseignées, les étudiants d'une université devraient consacrer en moyenne 45 heures de travail par semaine, avec un écart-type de 9 heures, selon la discipline. Désirant savoir si la situation actuelle diffère sensiblement ou non de cette opinion, on a considéré un échantillon aléatoire de 36 étudiants

inscrits l'année considérée dans cette université, auxquels il a été posé la question suivante : « **Combien d'heures par semaine consacrez-vous à vos études ?** »

1. Les réponses de ces 36 étudiants ayant donné une moyenne de 47.2 heures d'étude par semaine, peut-on affirmer, au seuil de signification de 5%, que le nombre d'heures d'étude des étudiants est significativement plus élevé que la norme ?
2. Calculer l'erreur de seconde espèce puis la puissance du test, en considérant précisément comme hypothèse alternative  $H_1 : \mu = 46$ .

#### Exercice 5 (4 points)

On a testé le pH, supposé gaussien, de deux types de solutions chimiques A et B. L'analyse de 6 échantillons de A a donné un pH moyen de 7.52 avec un écart-type de 0.024. L'analyse de 5 échantillons de B a donné un pH moyen de 7.49 avec un écart-type de 0.032. La différence entre les deux types de solutions est-elle significative au seuil de signification de 0.05 ?

#### Exercice 6 (bonus 2 points)

En se référant aux dates de début du pontificat (*dates de consécration*) et de fin (*par décès, démission ou inaptitude*), la durée d'exercice de chacun des 263 précédents papes (*excepté Jean-Paul II*) a été calculée en nombre d'années. Les résultats groupés en cinq tranches sont présentés dans le tableau suivant :

Pontificat	Nombre de papes
Moins d'une année	46
1 an – 5 ans	76
5 ans – 10 ans	67
10 ans - 20 ans	63
20 ans et plus	11
Total	263

Que penser, au seuil de signification de 5%, de l'hypothèse selon laquelle la distribution du pontificat des papes serait une distribution exponentielle ?

# Examen de Statistique

HEI 3 - 21 janvier 2013

- *Durée : 2 heures*
- *Autorisés : calculatrice et recto-verso manuscrit*
- *Le barème n'est donné qu'à titre indicatif*

## Exercice 1 (7,5 points).

Une école d'ingénieur souhaite créer un partenariat avec une grande entreprise. Avant d'accepter, le département des ressources humaines de l'entreprise décide de choisir au hasard 62 étudiants de dernière année et de leur faire passer un test permettant de mesurer diverses aptitudes. Nous présentons dans le tableau suivant, entre autres, la répartition des résultats  $X$  obtenus.

Classes $[a, b[$	Effectifs	$P(a \leq X < b)$
$60 \leq X < 80$	10	0,16
$80 \leq X < 90$	12	0,21
$90 \leq X < 100$	18	
$100 \leq X < 110$	15	0,20
$110 \leq X < 140$	7	0,16

1. L'échantillon a une moyenne de 94,8 et un écart-type de 15,7.  
Expliquer comment ont été calculées ces deux caractéristiques.
2. On veut savoir si  $X$  peut être supposée gaussienne.
  - (a) Préciser la loi de  $X$  utilisée pour calculer les probabilités dans le tableau ?
  - (b) Calculer  $P(90 \leq X < 100)$ .
  - (c) Est-il possible de supposer que  $X$  suive une loi normale ? La réponse sera justifiée à l'aide d'un test.
3. Donner un intervalle de confiance, au niveau 95%, pour :
  - (a) la moyenne théorique  $m$  de  $X$ .
  - (b) la variance théorique  $\sigma^2$  de  $X$ .
4. Que représentent  $m$  et  $\sigma^2$  ?
5. L'entreprise signera le partenariat à condition que l'école ait des résultats conformes à la moyenne nationale qui est, pour ce type de test, égale à 100.
  - (a) Que fera l'entreprise compte-tenu de l'échantillon observé ?
  - (b) Quelle est la probabilité que l'entreprise se trompe avec une telle décision ?

## Exercice 2 (5 points).

La distance d'arrêt d'un véhicule est égale à la distance de réaction (distance parcourue entre le moment où le conducteur perçoit un obstacle et celui où il appuie sur la pédale de frein) augmentée du chemin de freinage (distance parcourue entre le début du freinage et l'arrêt total). Le tableau suivant fournit pour 12 vitesses  $V$  un chemin de freinage  $C$  mesuré en mètres.

Vitesse $V$	Chemin de freinage $C$
40	8,8
50	11
60	19,8
70	27
80	39
90	44,6
100	58
110	78
120	79,4
130	93,1
140	108
150	124

- A l'aide de l'annexe 1, préciser pour chacun des modèles suivants une estimation (ponctuelle) des paramètres  $\alpha$  et  $\beta$  ainsi que du coefficient de détermination :
  - Modèle linéaire :  $C \simeq \alpha V + \beta$ .
  - Modèle puissance :  $C \simeq \alpha V^\beta$ .
- Pourquoi avons-nous utilisé le signe  $\simeq$  plutôt que le signe  $=$  ?
  - A quel modèle correspond le tableau inutilisé de l'annexe 1 ?
- Quel est le meilleur des modèles ?
  - Donner un intervalle de confiance, au niveau 95%, pour le  $\alpha$  du modèle choisi ?
  - A l'aide du modèle choisi, donner un intervalle de prévision pour le chemin de freinage à une vitesse de 170 km/h, au niveau 95%.

**Exercice 3** (2,5 points).

Une entreprise, oeuvrant dans le domaine de la transformation du métal en feuille, vient d'afficher divers postes dans un nouveau département. Avant d'en arriver à la sélection des candidats, la responsable des ressources humaines a administré un test de perception des formes à un échantillon d'individus provenant de trois secteurs de l'entreprise. Les résultats sont présentés dans le tableau suivant.

Atelier mécanique	Assemblage	Manutention
89	93	76
92	91	83
92	104	73
80	82	84
92	103	82
74	88	94
79	86	91
82	102	83
99	97	85
94	86	87
72	85	81
82	102	93
81	89	91
76	94	91
72	88	87
92	90	79

1. Calculer, en justifiant, les éléments manquants (cellules grisées) dans le tableau d'analyse de variance ci-dessous.

10	ANALYSE DE VARIANCE						
11	Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
12	Entre Groupes						
13	A l'intérieur des groupes	2429					
14							
15	Total	3095					
16							

2. Le secteur de l'entreprise a-t-il un effet significatif au seuil de 5% sur la performance (moyenne) au test ?

#### Exercice 4 (3,5 points).

Une entreprise fabrique une pièce mécanique particulière pour un fabricant d'automobiles. Quatre machines peuvent être utilisées pour usiner cette pièce. Cinq opérateurs sont susceptibles de travailler sur ces machines. Toutefois l'opération de ces machines requiert une certaine dextérité qui peut varier d'un opérateur à l'autre et affecter les résultats (le temps d'usinage). Les observations sont regroupées dans le tableau suivant.

	Machine 1	Machine 2	Machine 3	Machine 4
Opérateur 1	21,2	19,9	20,1	20,6
Opérateur 2	19,6	20,0	20,2	21,1
Opérateur 3	19,8	20,2	20,6	21,7
Opérateur 4	19,9	21,1	21,7	22,1
Opérateur 5	21,8	21,5	22,5	21,1

1. Indiquer si les opérateurs ont un effet significatif sur le temps d'usinage au seuil de 5% :  
(a) A l'aide du tableau obtenu avec une ANOVA à 1 facteur :

11	ANALYSE DE VARIANCE						
12	Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
13	Entre Groupes	6,053	4	1,51325	2,825863679	0,062547376	3,055568276
14	A l'intérieur des groupes	8,0325	15	0,5355			
15							
16	Total	14,0855	19				
17							

- (b) A l'aide du tableau obtenu cette fois avec une ANOVA à 2 facteurs que l'on complètera au besoin :

15	ANALYSE DE VARIANCE						
16	Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
17	Lignes					0,04939031	
18	Colonnes	2,4855					
19	Erreur	5,547					
20							
21	Total						
22							

2. Comment peut-on justifier cette contradiction apparente ?
3. Les machines ont-elles un effet significatif sur le temps d'usinage au seuil de 5% ?

#### Exercice 5 (1,5 points).

Une directrice marketing veut étudier l'effet de deux facteurs sur les ventes d'une denrée alimentaire et ceci pour une période de quinze jours. Les facteurs considérés sont la hauteur de la tablette et la position de l'allée. Dix-huit magasins ont été sélectionnés et 2 ont été affectés à chaque traitement (combinaison position de l'allée et hauteur de la tablette). Le nombre d'unités vendues pendant la quinzaine observée est présenté dans le tableau suivant.

	45 cm	75 cm	115 cm
À l'avant	70	82	106
	60	74	92
Au milieu	38	40	52
	36	34	44
À l'arrière	48	60	80
	44	48	66

Les questions suivantes sont à traiter au seuil de 0,1% à l'aide de l'annexe 2.

1. L'interaction des deux facteurs a-t-elle un effet significatif sur les ventes ?
2. La hauteur de la tablette a-t-elle un effet significatif sur les ventes ?
3. (a) La position de l'allée a-t-elle un effet significatif sur les ventes ?  
(b) Est-il possible de privilégier une position de l'allée par rapport aux deux autres ?

## Annexe 1 : trois régressions linéaires

Régression linéaire entre V (variable explicative) et C (variable expliquée)

	A	B	C	D	E	F
1	RAPPORT DÉTAILLÉ					
2						
3	<i>Statistiques de la régression</i>					
4	Coefficient de détermination multiple	0,990374008				
5	Coefficient de détermination R^2	0,980840675				
6	Coefficient de détermination R^2	0,978924743				
7	Erreur-type	5,628335427				
8	Observations	12				
9						
10	ANALYSE DE VARIANCE					
11		<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
12	Régression	1	16217,28757	16217,28757	511,939069	6,40468E-10
13	Résidus	10	316,7815967	31,67815967		
14	Total	11	16534,06917			
15						
16		<i>Coefficients</i>	<i>Erreur-type</i>	<i>Statistique t</i>	<i>Probabilité</i>	
17	Constante	-43,61002331	4,757366007	-9,166842165	3,50703E-06	
18	V	1,06493007	0,047066505	22,62607056	6,40468E-10	
19						

Régression linéaire entre V (variable explicative) et  $Y = \ln C$  (variable expliquée)

	A	B	C	D	E	F
1	RAPPORT DÉTAILLÉ					
2						
3	<i>Statistiques de la régression</i>					
4	Coefficient de détermination multiple	0,973320942				
5	Coefficient de détermination R^2	0,947353656				
6	Coefficient de détermination R^2	0,942089022				
7	Erreur-type	0,212670701				
8	Observations	12				
9						
10	ANALYSE DE VARIANCE					
11		<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
12	Régression	1	8,138778764	8,138778764	179,9467132	1,01787E-07
13	Résidus	10	0,45228827	0,045228827		
14	Total	11	8,591067035			
15						
16		<i>Coefficients</i>	<i>Erreur-type</i>	<i>Statistique t</i>	<i>Probabilité</i>	
17	Constante	1,49554027	0,179760495	8,319627003	8,34123E-06	
18	V	0,023856768	0,001778442	13,41442184	1,01787E-07	
19						

Régression linéaire entre  $X = \ln V$  (variable explicative) et  $Y = \ln C$  (variable expliquée)

	A	B	C	D	E	F
1	RAPPORT DÉTAILLÉ					
2						
3	<i>Statistiques de la régression</i>					
4	Coefficient de détermination multiple	0,995671732				
5	Coefficient de détermination R^2	0,991362198				
6	Coefficient de détermination R^2	0,990498418				
7	Erreur-type	0,086144029				
8	Observations	12				
9						
10	ANALYSE DE VARIANCE					
11		<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
12	Régression	1	8,516859097	8,516859097	1147,701897	1,18764E-11
13	Résidus	10	0,074207938	0,007420794		
14	Total	11	8,591067035			
15						
16		<i>Coefficients</i>	<i>Erreur-type</i>	<i>Statistique t</i>	<i>Probabilité</i>	
17	Constante	-5,56253445	0,276359804	-20,12787089	2,01641E-09	
18	X = ln V	2,082185828	0,061461752	33,8777493	1,18764E-11	
19						

## Annexe 2

	A	B	C	D	E	F	G
1	Analyse de variance: deux facteurs avec répétition d'expérience						
2							
3	RAPPORT DÉTAILLÉ	45 cm	75 cm	115 cm	Total		
4	<i>À l'avant</i>						
5	Nombre d'échantillons	2	2	2	6		
6	Somme	130	156	198	484		
7	Moyenne	65	78	99	80,66666667		
8	Variance	50	32	98	271,4666667		
9							
10	<i>Au milieu</i>						
11	Nombre d'échantillons	2	2	2	6		
12	Somme	74	74	96	244		
13	Moyenne	37	37	48	40,66666667		
14	Variance	2	18	32	42,66666667		
15							
16	<i>À l'arrière</i>						
17	Nombre d'échantillons	2	2	2	6		
18	Somme	92	108	146	346		
19	Moyenne	46	54	73	57,66666667		
20	Variance	8	72	98	189,4666667		
21							
22	<i>Total</i>						
23	Nombre d'échantillons	6	6	6			
24	Somme	296	338	440			
25	Moyenne	49,33333333	56,33333333	73,33333333			
26	Variance	175,4666667	363,8666667	565,8666667			
27							
28							
29	ANALYSE DE VARIANCE						
30	<i>Source des variations</i>	<i>Somme des carrés</i>	<i>Degré de liberté</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Probabilité</i>	<i>Valeur critique pour F</i>
31	Échantillon	4836	2	2418	53,07804878	1,04304E-05	
32	Colonnes	1828	2	914	20,06341463	0,000482122	
33	Interaction	280	4	70	1,536585366	0,271576363	
34	A l'intérieur du groupe	410	9	45,55555556			
35							

# Examen de Probabilités et Statistique

HEI3

03 Avril 2013

- *Durée : 2 heures*
- *Autorisés : calculatrice et recto-verso manuscrit*
- *Tables statistiques en annexe*

## **Exercice 1.**

On mesure la force de compression d'un ciment en moulant de petits cylindres et en mesurant la pression  $X$  (exprimée en  $\text{kg/cm}^2$ ) à partir de laquelle ils se cassent. Pour 10 cylindres utilisés, on relève les pressions suivantes :

19.6 19.9 20.4 19.8 20.5 21.0 18.5 19.7 18.4 19.4

On suppose que  $X$  suit une loi de Gauss de moyenne  $m$  et d'écart-type  $\sigma$ . (**Les intervalles de confiance seront déterminés au niveau de 95%**).

- Admettons que  $\sum_{i=1}^{10} x_i = 197.2$  et  $\sum_{i=1}^{10} x_i^2 = 3894.88$ .
  - Déterminer une estimation ponctuelle non biaisée pour  $m$  et  $\sigma^2$ .
  - Déterminer maintenant un intervalle de confiance pour  $m$  et  $\sigma^2$ .
- Supposons dans cette question  $\sigma^2 = 0,69$ .
  - Déterminer dans ces conditions un intervalle de confiance pour  $m$ .
  - Comparer le résultat du 2.a) avec celui du 1.b).
  - Déterminer la taille minimale de l'échantillon pour avoir une estimation à  $\pm 0.2 \text{ kg/cm}^2$

## **Exercice 2.**

Sur 4000 naissances, on relève 2065 garçons.

- Que pouvons-nous alors penser de l'idée selon laquelle il y'a une chance sur deux d'avoir un garçon à la naissance (On répondra à la question successivement au seuil de 5% puis 1%) ?
- Au seuil de 5 %, quelle est la probabilité de se tromper en suivant la conclusion du test ?

## **Exercice 3.**

Dans une école d'ingénieurs, les notes de Statistique de deux promotions suivent des lois normales :  $X_1 \sim N(m_1; \sigma_1)$  et  $X_2 \sim N(m_2; \sigma_2)$ . Un échantillon de 25 notes pour  $X_1$  fournit une moyenne de 12 et un écart-type de 3. Quant à  $X_2$ , un échantillon de 10 notes fournit une moyenne de 13 et un écart-type de 4. Peut-on considérer au seuil de 5% que la deuxième promotion est meilleure que la première ?

#### **Exercice 4.**

Chez un individu adulte, le logarithme du dosage de d-dimères, variable que nous noterons  $X$ , est modélisé par une loi normale d'espérance  $\mu$  et de variance  $\sigma^2$ . La variable  $X$  est un indicateur de risque cardio-vasculaire : on considère que chez les individus sains,  $\mu$  vaut  $-1$ , alors que chez les individus à risque,  $\mu$  vaut  $0$ . On souhaite étudier l'influence de la consommation d'huile d'olive sur le risque cardio-vasculaire.

1. On a fait suivre un régime à base d'huile d'olive à un groupe de 13 patients, précédemment considérés comme à risque. Après le régime, on a mesuré la valeur de  $X$  pour chaque patient, et obtenu une moyenne empirique de  $-0.15$ . On suppose  $\sigma^2$  connu et égal à  $0.09$ .
  - a) Donner la règle de décision du test de  $H_0 : \mu = 0$ , contre  $H_1 : \mu = -1$ , au seuil de 5%.
  - b) Quelle est votre conclusion ?
  - c) Calculer le risque de deuxième espèce et la puissance du test.
2. Pour le même groupe de 13 patients, on a observé un écart-type empirique égal à  $0.37$ .
  - a) Donner la règle de décision du test de  $H_0 : \sigma^2 = 0.09$ , contre  $H_1 : \sigma^2 \neq 0.09$ , au seuil de 5%
  - b) Quelle est votre conclusion ?
3. En supposant la variance inconnue et en utilisant l'estimation de la question précédente.
  - a) Donner la règle de décision du test de  $H_0 : \mu = 0$ , contre  $H_1 : \mu < 0$ , au seuil de 5%.
  - b) Quelle est votre conclusion ?
4. On suppose dans cette question que  $X$  n'est plus supposée gaussienne. On reprend l'expérience sur un groupe de 130 patients, pour lesquels on observe une moyenne empirique de  $-0.12$  et un écart-type de  $0.32$ .
  - a) Donner la règle de décision du test de  $H_0 : \mu = 0$ , contre  $H_1 : \mu < 0$ , au seuil de 5%.
  - b) Quelle est la  $p$ -valeur correspondant à  $-0.12$  ? (la  $p$ -valeur est le seuil pour lequel la valeur observée de la statistique de test serait la limite de la région de rejet. C'est la probabilité sous  $H_0$  que la statistique de test soit au-delà de la valeur déjà observée.)
  - c) Quelle est votre conclusion ?

# Examen de Statistique

HEI 3 - 11 juin 2013

- *Durée : 2 heures*
- *Autorisés : calculatrice et recto-verso manuscrit*
- *Le barème n'est donné qu'à titre indicatif*

**Sauf mention contraire, on réalisera les tests au seuil de 5% en détaillant bien les étapes.**

## Exercice 1 (5 points).

Le tableau ci dessous indique pour un échantillon de 9 étudiants de troisième année, le nombre X de jours consacrés à préparer les examens de fin de semestre et Y la moyenne obtenue sur 20.

X	43	45	47	53	55	57	61	65	67
Y	7,2	7,3	8,1	8,4	10	10,4	12	12,6	14,3

**Les options graphiques d'Excel et son utilitaire d'analyse permettent d'obtenir les résultats fournis en Annexe.**

- Pour le modèle linéaire, donner l'ordonnée à l'origine de la droite.
  - Pour le modèle exponentiel, donner le coefficient de détermination.
  - Pour le modèle puissance, compléter le résultat fourni par l'utilitaire d'analyse.
- Quel modèle ajuste au mieux le nuage de points ?
- En utilisant le meilleur des modèles, donner pour un étudiant ayant consacré 75 jours à préparer ses examens :
  - Une estimation ponctuelle de sa moyenne.
  - Un intervalle de prévision de sa moyenne.

## Exercice 2 (4 points).

- En réalisant un test de comparaison à partir des données suivantes, peut-on rejeter l'idée selon laquelle il y'aurait plus de femmes à fumer que d'hommes ?

	Homme	Femme
Fumeur	24	41
Non fumeur	23	35

- En réalisant un test du chi 2, étudier la dépendance entre l'environnement familial et le fait de fumer.

	Père fumeur et mère fumeuse	Père fumeur et mère non fumeuse	Père non fumeur et mère fumeuse	Père non fumeur et mère non fumeuse
Fumeur	13	16	7	29
Non fumeur	5	24	6	23

**Exercice 3** (6 points).

Dans le cadre d'un concours comptant 1500 candidats, un des correcteurs rend ses 100 copies au secrétariat qui établit la distribution suivante :

Classes	[0,4[	[4,8[	[8,12[	[12,16[	[16,20[
Effectifs	8	22	30	28	12

1. Peut-on considérer que les notes suivent une loi normale ? La réponse sera justifiée par un test.
2. Estimer par intervalle de confiance, au niveau 95%, la moyenne et la variance des candidats.
3. Calculer la barre d'admission si le concours offre 150 places.

**Exercice 4** (5 points).

Dans cet exercice, on utilisera les sorties Excel fournies que l'on complétera au besoin. Le détail des tests mis en oeuvre pour répondre n'est pas attendu.

Nous nous proposons d'analyser l'influence du temps et de trois espèces ligneuses d'arbre sur la décomposition de la masse d'une litière constituée de feuilles de Lierre.

1. Pour ce faire, 12 sachets d'une masse identique de feuilles de lierre ont été constitués, sachets permettant une décomposition naturelle. Puis une première série de 4 sachets, choisis au hasard, a été déposée sous un chêne, une deuxième sous un peuplier, et la dernière série sous un frêne. Après 2, 7, 10 et 16 semaines respectivement, un sachet est prélevé au hasard sous chaque arbre et la masse résiduelle est déterminée pour chacun d'eux. Cette masse est exprimée en pourcentage de la masse initiale.

Semaine	Chêne	Peuplier	Frêne
2	85	85	84
7	75	73	72
10	71	74	67
16	62	67	58

Une ANOVA à deux facteurs sans répétition fournit le résultat suivant :

	A	B	C	D	E	F	G
1	Analyse de variance: deux facteurs sans répétition d'expérience						
2							
3	RAPPORT DÉTAILLÉ	Nombre d'échantillons	Somme	Moyenne	Variance		
4	2	3	254	84,66666667	0,333333333		
5	7	3	220	73,33333333	2,333333333		
6	10	3	212	70,66666667	12,33333333		
7	16	3	187	62,33333333	20,33333333		
8							
9	Chêne	4	293	73,25	90,91666667		
10	Peuplier	4	299	74,75	56,25		
11	Frêne	4	281	70,25	117,5833333		
12							
13							
14	ANALYSE DE VARIANCE						
15	Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
16	Lignes						
17	Colonnes						
18	Erreur	28,66666667					
19							
20	Total	836,25					

- (a) Le temps a-t-il un effet significatif sur la décomposition ?
  - (b) L'espèce de l'arbre a-t-elle un effet significatif sur la décomposition ?
2. En utilisant 24 sachets, nous disposons maintenant de deux observations par arbre pour chaque semaine.

Semaine	Chêne	Peuplier	Frêne
2	85	85	84
	87	84	85
7	75	73	72
	72	75	70
10	71	74	67
	66	71	64
16	62	67	58
	64	60	59

Une ANOVA à deux facteurs avec répétitions fournit le résultat suivant :

35	ANALYSE DE VARIANCE						
36	<i>Source des variations</i>	<i>Somme des carrés</i>	<i>Degré de liberté</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Probabilité</i>	<i>Valeur critique pour F</i>
37	Échantillon	1718,833333					
38	Colonnes	61,58333333					
39	Interaction					0,443034267	
40	A l'intérieur du groupe	60					
41							
42	Total						

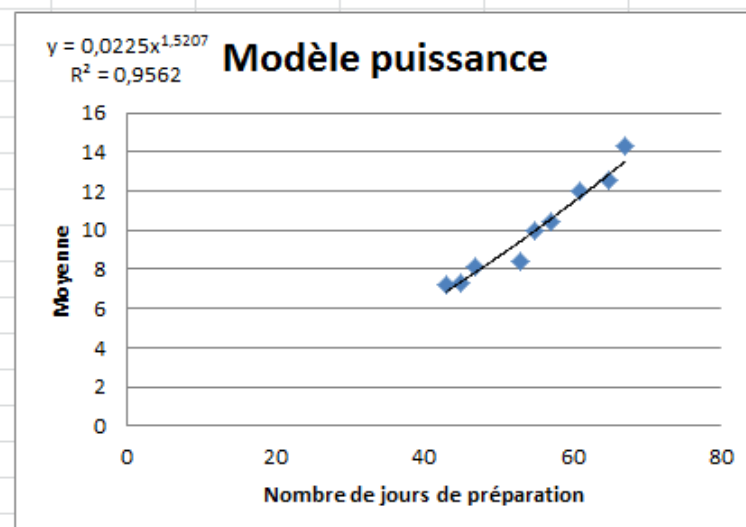
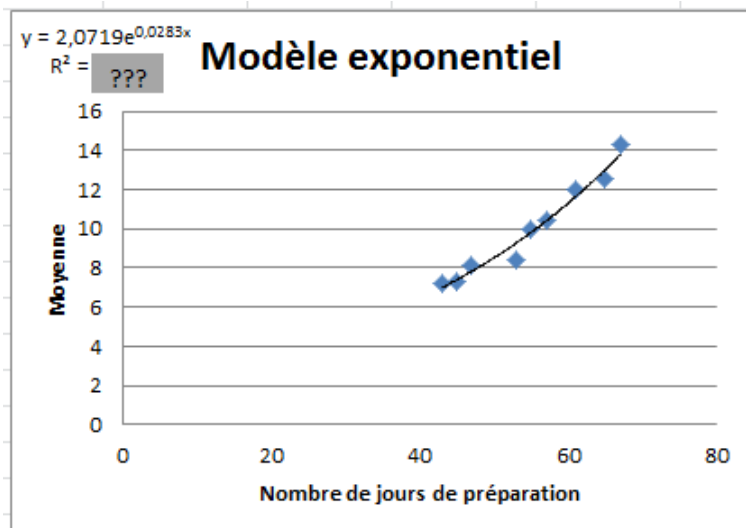
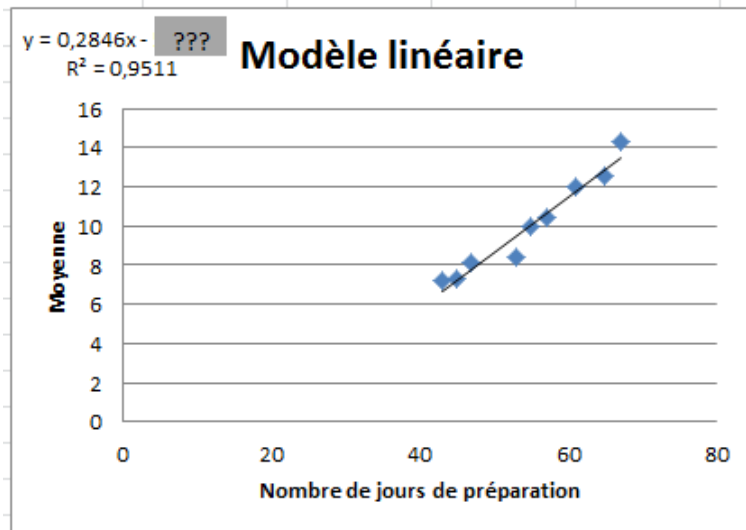
- L'interaction entre le temps et l'espèce de l'arbre a-t-elle un effet significatif sur la décomposition ?
- Le temps a-t-il un effet significatif sur la décomposition ?
- L'espèce de l'arbre a-t-elle un effet significatif sur la décomposition ?
- L'ANOVA à deux facteurs fournissant des réponses différentes suivant qu'elle est sans ou avec répétitions, quelle réponse doit-on privilégier quant à l'effet de l'espèce de l'arbre sur la décomposition ?

3. En ne considérant que les différentes espèces, une ANOVA à un facteur fournit le résultat suivant :

10	ANALYSE DE VARIANCE						
11	<i>Source des variations</i>	<i>Somme des carrés</i>	<i>Degré de liberté</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Probabilité</i>	<i>Valeur critique pour F</i>
12	Entre Groupes	61,58333333	2	30,79166667	0,357202044	0,703800357	3,466800112
13	A l'intérieur des groupes	1810,25	21	86,20238095			
14							
15	Total	1871,833333	23				

- L'espèce de l'arbre a-t-elle un effet significatif sur la décomposition ?
- Donner la décomposition des carrés totaux à l'origine de l'ANOVA à un facteur.
- Indiquer la variable de décision utilisée dans l'ANOVA à un facteur ainsi que sa loi.
- Expliquer pourquoi la conclusion n'est pas pertinente en comparaison avec l'ANOVA à deux facteurs.

## Annexe 1 : trois graphiques



## Annexe 2 : trois régressions linéaires

Régression linéaire entre X (variable explicative) et Y (variable expliquée)

	A	B	C
1	RAPPORT DÉTAILLÉ		
2			
3	<i>Statistiques de la régression</i>		
4	Coefficient de détermination multiple	0,975257205	
5	Coefficient de détermination R^2	0,951126615	
6	Coefficient de détermination R^2	0,944144703	
7	Erreur-type	0,594964656	
8	Observations	9	
9			
10	ANALYSE DE VARIANCE		
11		<i>Degré de liberté</i>	<i>Somme des carrés</i>
12	Régression	1	48,2221194
13	Résidus	7	2,477880597
14	Total	8	50,7
15			
16		<i>Coefficients</i>	<i>Erreur-type</i>
17	Constante	???	1,350116302
18	X	0,284552239	0,024379794

Régression linéaire entre X (variable explicative) et ln Y (variable expliquée)

	A	B	C
1	RAPPORT DÉTAILLÉ		
2			
3	<i>Statistiques de la régression</i>		
4	Coefficient de détermination multiple	0,983806065	
5	Coefficient de détermination R^2	0,967874374	
6	Coefficient de détermination R^2	0,963284999	
7	Erreur-type	0,047547323	
8	Observations	9	
9			
10	ANALYSE DE VARIANCE		
11		<i>Degré de liberté</i>	<i>Somme des carrés</i>
12	Régression	1	0,476779501
13	Résidus	7	0,015825235
14	Total	8	0,492604736
15			
16		<i>Coefficients</i>	<i>Erreur-type</i>
17	Constante	0,728473013	0,107896183
18	X	0,028294215	0,001948341

Régression linéaire entre ln X (variable explicative) et ln Y (variable expliquée)

	A	B	C
1	RAPPORT DÉTAILLÉ		
2			
3	<i>Statistiques de la régression</i>		
4	Coefficient de détermination multiple	0,977836092	
5	Coefficient de détermination R^2	0,956163422	
6	Coefficient de détermination R^2	0,949901054	
7	Erreur-type	0,055541627	
8	Observations	9	
9			
10	ANALYSE DE VARIANCE		
11		<i>Degré de liberté</i>	<i>Somme des carrés</i>
12	Régression	1	0,47101063
13	Résidus	7	0,021594106
14	Total	8	0,492604736
15			
16		<i>Coefficients</i>	<i>Erreur-type</i>
17	Constante	???	0,491635073
18	ln X	???	0,123065531

# Examen de Statistique

HEI 3 - novembre 2013

- *Durée : 2 heures*
- *Autorisés : calculatrice et recto-verso manuscrit*
- *Le barème n'est donné qu'à titre indicatif*

**Sauf mention contraire, on réalisera les tests au seuil de 5% en détaillant bien les étapes.**

## Exercice 1 (4 points).

Une entreprise remplit des boites de 200 g de chocolats.

1. On suppose dans cette question (uniquement) que le poids d'une boite de chocolats suive une loi normale. La probabilité qu'il soit supérieur à 230 g est de 0,02 et celle qu'il soit inférieur à 180 g est de 0,015.
  - (a) Déterminer les paramètres de la loi.
  - (b) Quelle est la probabilité que le poids d'une boite soit inférieur à 200 g ?
  - (c) Quelle est la probabilité que la différence entre les poids de deux boites soit inférieure à 20g ?
2. On suppose maintenant que les boites de chocolats aient un poids moyen de 207 g avec un écart type de 20 g. L'entreprise livre les chocolats aux grandes surfaces par cartons de 100 boites.
  - (a) Quelle est la loi du poids moyen de 100 boites ?
  - (b) Quelle est la probabilité que la différence entre les poids moyens de 100 boites dans deux cartons soit supérieure à 10 g ?

## Exercice 2 (5 points).

Sur une parcelle de soja, on a mesuré la hauteur en cm de 100 plantes à l'âge de 6 semaines.

On obtient les résultats suivants après regroupement par classes et en considérant pour les hauteurs les centres des classes :

Hauteurs	36	37	38	39	40	41
Effectifs	6	11	26	32	14	11

Supposons la hauteur gaussienne de moyenne  $m$  et d'écart-type  $\sigma$ .

1.
  - (a) Déterminer un intervalle de confiance à 95% pour  $\sigma$ .
  - (b) Déterminer un intervalle de confiance à 95% pour  $m$ .
2. Supposons maintenant  $\sigma = 2$ .
  - (a) Déterminer un intervalle de confiance à 95% pour  $m$ . Calculer son incertitude.
  - (b) Comparer et commenter les deux intervalles obtenus pour  $m$ .
  - (c) Déterminer le nombre de plantes à mesurer pour avoir une estimation de  $m$  avec une confiance de 95% et une incertitude n'excédant pas 2mm.

**Exercice 3** (5 points).

On étudie le taux de réussite au baccalauréat 2013 dans une série donnée.

1. Le taux de réussite nationale au baccalauréat 2012 dans cette série était de 67%.
  - (a) Dans un centre d'examen A comptant 3500 candidats, il y a eu 216 reçus sur 300 candidats choisis au hasard sans remise.
    - i. L'échantillon peut-il être considéré comme étant aléatoire simple autrement dit le tirage peut-il être considéré comme étant avec remise ?
    - ii. Les résultats de ce centre sont ils conformes aux résultats nationaux de 2012 ?
  - (b) Dans un centre d'examen B comptant 2500 candidats, il y a eu 128 reçus sur 200 candidats choisis au hasard sans remise.

Les résultats de ces deux centres sont ils significativement différents ?
2. On souhaite maintenant estimer, avec une confiance de 98% et une incertitude n'excédant pas 2%, le taux de réussite nationale au baccalauréat 2013 dans cette série. Déterminer alors la taille minimale de l'échantillon à observer ?

**Exercice 4** (6 points).

On souhaite interpréter l'influence de la prise d'un médicament sur un taux d'anticorps produit.

L'échantillon A, constitué d'individus non traités, est de taille 50. On observe un taux moyen de 1,6 et une variance de 1,8.

L'échantillon B, constitué d'individus traités, est de taille 40. On observe un taux moyen de 3 et une variance de 2,3.

1. En supposant le taux d'anticorps produit gaussien, le traitement est-il efficace ?
2. En ne supposant plus le taux d'anticorps produit gaussien, le traitement est-il efficace ?

# Examen de Statistique

HEI 3 - 20 janvier 2014

- *Durée : 2 heures*
- *Autorisés : calculatrice et recto-verso manuscrit*
- *Le barème n'est donné qu'à titre indicatif*

**Sauf mention contraire, on réalisera les tests au seuil de 5% en détaillant bien toutes les étapes : hypothèses, variable de décision et sa loi, région critique, valeur observée de la variable de décision et enfin conclusion**

## Exercice 1 (5 points).

En biostatistique, une étude de la croissance d'une population de campagnols des champs (*Microtus Arvalis*) a fourni, pour le nombre d'individus  $N$  suivant le temps  $T$  (en mois), les résultats suivants :

<b>T</b>	2	4	6	8	10	12	14
<b>N</b>	5	16	20	40	100	200	320

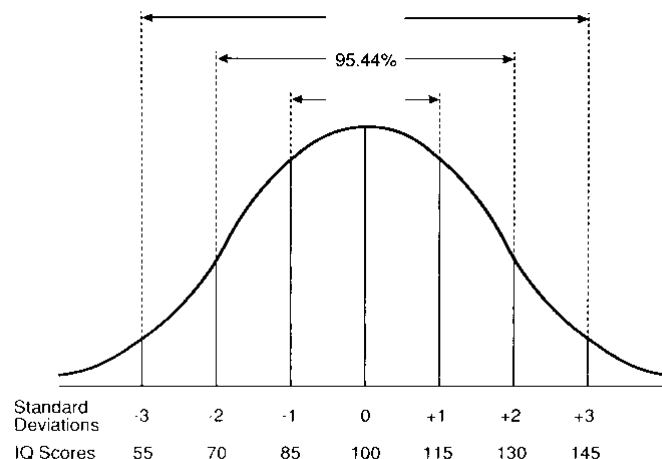
**En Annexe, vous trouverez les résultats fournis par les options graphiques d'Excel et son utilitaire d'analyse.**

- (a) Pour le modèle linéaire, donner le coefficient directeur de la droite.  
(b) Pour le modèle puissance, donner le coefficient de détermination.  
(c) Pour le modèle exponentiel, compléter le résultat fourni par l'utilitaire d'analyse.
- Quel modèle ajuste au mieux le nuage de points ?
- En utilisant le meilleur des modèles, donner pour 18 mois :
  - Une estimation ponctuelle du nombre d'individus.
  - Un intervalle de prévision du nombre d'individus.

## Exercice 2 (5,5 points).

- Une étude menée en 1991 a mis en évidence le caractère gaussien du QI au sein de la population :

Figure 1  
The Normal Distribution of IQ Scores



- (a) A partir de la courbe de Gauss ci-dessus, déterminer le QI moyen.
- (b) En utilisant l'aire sous la courbe entre 70 et 130, justifier que l'écart-type du QI vaut 15.
- (c) Déterminer l'aire sous la courbe :
  - à gauche de 85 c'est à dire  $P(QI \leq 85)$
  - entre 85 et 100
  - entre 100 et 115
  - à droite de 115

2. Une étude plus récente sur 40 individus fournit les résultats suivants :

QI	Effectifs
< 85	1
[85 ; 100[	11
[100 ; 115[	19
$\geq 115$	9

La modélisation du QI de 1991, par la loi normale de moyenne 100 et d'écart-type 15, est-elle encore adaptée ? La réponse sera justifiée à l'aide d'un test du chi 2.

### Exercice 3 (2 points).

On désire savoir s'il existe une dépendance entre l'âge d'un individu atteint du cancer et ses chances de guérison. On mène une enquête sur trois classes d'âge et on obtient les résultats suivants :

	Guéri	Non guéri
50 - 60 ans	1409	507
60 - 70 ans	763	248
70 - 80 ans	571	192

Quelle est votre conclusion ? La réponse sera justifiée à l'aide d'un test du chi 2.

### Exercice 4 (7,5 points).

**Dans cet exercice, on utilisera les sorties Excel fournies que l'on complétera au besoin. Les différentes étapes des tests mis en oeuvre pour répondre aux questions ne sont pas attendues ici.**

Dans le cadre d'une étude pédagogique, quatre méthodes d'apprentissage sont expérimentées.

1. A l'issue de la formation, une évaluation est proposée à 12 étudiants (trois pour chacune des méthodes M1, M2, M3 et M4) dont les résultats sont regroupés dans le tableau suivant :

M1	M2	M3	M4
15	45	56	39
27	65	47	43
37	76	43	72

Une ANOVA à un facteur fournit :

	A	B	C	D	E	F	G
1	Analyse de variance: un facteur						
2							
3	RAPPORT DÉTAILLÉ						
4	Groupes	Nombre d'échantillons	Somme	Moyenne	Variance		
5	M1	3	79	26,33333333	121,3333333		
6	M2	3	186	62	247		
7	M3	3	146	48,66666667	44,33333333		
8	M4	3	154	51,33333333	324,3333333		
9							
10							
11	ANALYSE DE VARIANCE						
12	Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
13	Entre Groupes						
14	A l'intérieur des groupes						
15							
16	Total						

La méthode d'apprentissage a-t-elle un effet significatif sur la note obtenue lors de l'évaluation ?

2. En considérant le QI des étudiants (C1 si  $QI < 100$ , C2 si  $100 \leq QI < 115$  et C3 si  $QI \geq 115$ ), on peut en fait voir les données précédentes de la manière suivante :

	M1	M2	M3	M4
C1	15	45	56	39
C2	27	65	47	43
C3	37	76	43	72

Une ANOVA à deux facteurs sans répétition fournit :

	A	B	C	D	E	F	G
1	Analyse de variance: deux facteurs sans répétition d'expérience						
2							
3	RAPPORT DÉTAILLÉ	Nombre d'échantillons	Somme	Moyenne	Variance		
4	C1	4	155	38,75	300,25		
5	C2	4	182	45,5	243,6666667		
6	C3	4	228	57	394		
7							
8	M1	3	79	26,33333333	121,3333333		
9	M2	3	186	62	247		
10	M3	3	146	48,66666667	44,33333333		
11	M4	3	154	51,33333333	324,3333333		
12							
13							
14	ANALYSE DE VARIANCE						
15	Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
16	Lignes					0,155615643	
17	Colonnes						
18	Erreur	792,8333333					
19							
20	Total						

(a) La méthode d'apprentissage a-t-elle un effet significatif sur la note ?

(b) Commenter les réponses obtenues aux questions 1. et 2.(a).

(c) Le QI a-t-il un effet significatif sur la note ?

3. En choisissant, par couple  $(C, M)$ , deux étudiants plutôt qu'un seul, on obtient les données suivantes :

	M1	M2	M3	M4
C1	15	45	56	39
	20	45	54	26
C2	27	65	47	43
	42	95	58	28
C3	37	76	43	72
	67	78	82	65

Une ANOVA à deux facteurs avec répétition fournit :

	A	B	C	D	E	F	G
1	Analyse de variance: deux facteurs avec répétition d'expérience						
2							
29	ANALYSE DE VARIANCE						
30	Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
31	Échantillon	3027,083333	2	1513,541667	8,76780111	0,004497928	3,885293835
32	Colonnes	3575,458333	3	1191,819444	6,904095261	0,005913904	3,490294819
33	Interaction						
34	A l'intérieur du groupe	2071,5	12	172,625			
35							
36	Total	10046,95833	23				

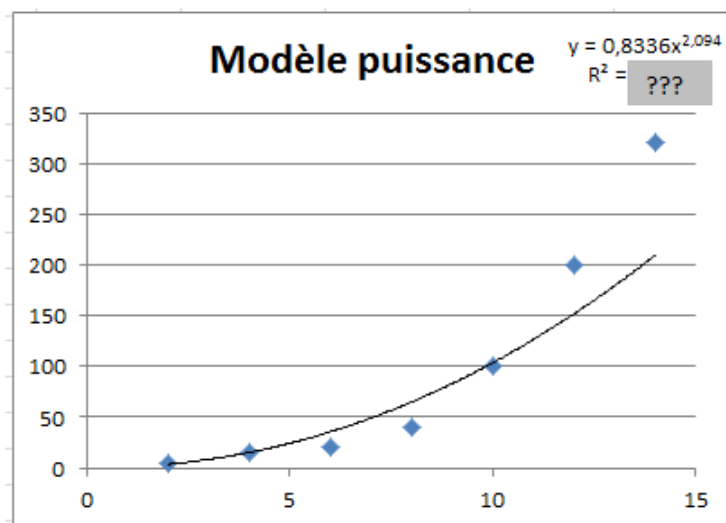
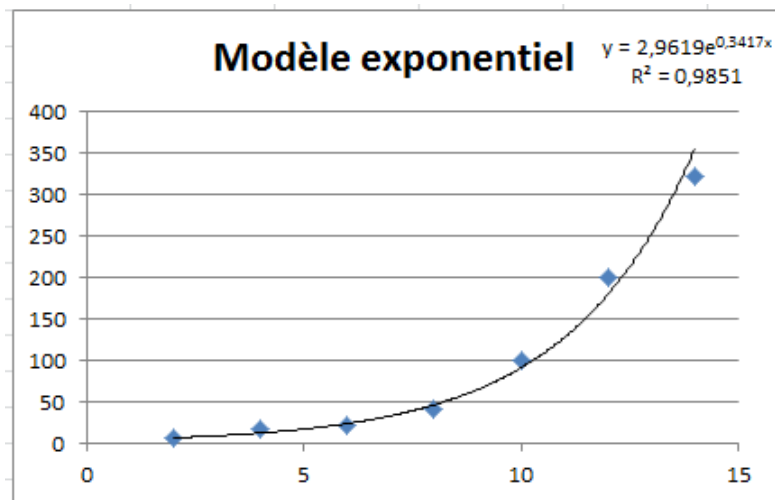
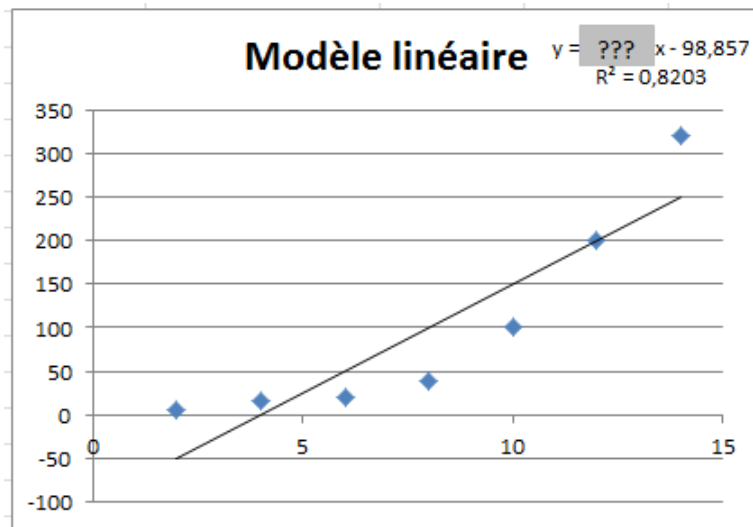
(a) L'interaction entre la méthode d'apprentissage et le QI a-t-elle un effet significatif sur la note ?

(b) Le QI a-t-il un effet significatif sur la note ?

(c) Commenter les réponses obtenues aux questions 2.(c) et 3.(b).

(d) Est-il possible de considérer que la méthode 2 soit plus efficace que la méthode 3 ?

## Annexe 1 : trois graphiques



## Annexe 2 : trois régressions linéaires

Régression linéaire entre T (variable explicative) et N (variable expliquée)

	A	B	C
1	RAPPORT DÉTAILLÉ		
2			
3	Statistiques de la régression		
4	Coefficient de détermination multiple	0,905717888	
5	Coefficient de détermination R^2	0,820324892	
6	Coefficient de détermination R^2	0,78438987	
7	Erreur-type	55,0982888	
8	Observations	7	
9			
10	ANALYSE DE VARIANCE		
11		Degré de liberté	Somme des carrés
12	Régression	1	69301,75
13	Résidus	5	15179,10714
14	Total	6	84480,85714
15			
16		Coefficients	Erreur-type
17	Constante		46,56655321
18	T		5,206298922

Régression linéaire entre T (variable explicative) et ln N (variable expliquée)

	A	B	C
1	RAPPORT DÉTAILLÉ		
2			
3	Statistiques de la régression		
4	Coefficient de détermination multiple	0,99251865	
5	Coefficient de détermination R^2	0,98509327	
6	Coefficient de détermination R^2	0,982111924	
7	Erreur-type	0,198964094	
8	Observations	7	
9			
10	ANALYSE DE VARIANCE		
11		Degré de liberté	Somme des carrés
12	Régression	1	13,08020027
13	Résidus	5	0,197933553
14	Total	6	13,27813383
15			
16		Coefficients	Erreur-type
17	Constante	???	0,168155351
18	T	???	0,01880034

Régression linéaire entre ln T (variable explicative) et ln N (variable expliquée)

	A	B	C
1	RAPPORT DÉTAILLÉ		
2			
3	Statistiques de la régression		
4	Coefficient de détermination multiple	0,963973713	
5	Coefficient de détermination R^2	0,929245318	
6	Coefficient de détermination R^2	0,915094382	
7	Erreur-type	0,433472059	
8	Observations	7	
9			
10	ANALYSE DE VARIANCE		
11		Degré de liberté	Somme des carrés
12	Régression	1	12,33864369
13	Résidus	5	0,939490131
14	Total	6	13,27813383
15			
16		Coefficients	Erreur-type
17	Constante	-0,181975408	0,520298897
18	ln(T)	2,094032555	0,258410899

## Examen de Statistique

HEI 3 – Mars 2014

- Durée : 2 heures
- Autorisés : calculatrice et recto-verso manuscrit
- Le barème n'est donné qu'à titre indicatif

(Sauf mention contraire, les tests seront réalisés au seuil de 5%, en détaillant scrupuleusement les différentes étapes)

### Exercice 1. (6 points)

La durée de vie d'un appareil d'un certain type est modélisée par une variable aléatoire normale de moyenne et d'écart-type inconnus. Comme on voudrait, cependant, que les  $\frac{3}{4}$  de la production de ces appareils aient une durée de vie entre 100 et 180 jours, et que par ailleurs, 3% de cette production ait une durée de vie supérieure à 180 jours :

- 1) Déterminer les paramètres  $m$  et  $\sigma$  de la durée de vie d'un appareil
- 2) Quelle est la probabilité que la durée de vie d'un appareil soit comprise entre 80 et 100 jours.
- 3) Déterminer la durée de vie maximale prévisible des 30% des appareils qui durent le moins.
- 4) Déterminer la durée de vie minimale prévisible des 20% des appareils qui durent le plus.
- 5) On considère un deuxième type d'appareil dont la durée de vie est cette fois-ci modélisée par une variable aléatoire normale de moyenne 110 et d'écart-type 25. Les durées de vie des deux types d'appareils étant supposées indépendantes, déterminer la probabilité qu'elles s'écartent l'une de l'autre, de moins de 10 jours.

### Exercice 2. (7 points)

Les caractéristiques les plus importantes du dispositif pyrotechnique qui permet d'assurer le gonflement d'un airbag sont la moyenne  $m$  et l'écart-type  $\sigma$  du délai entre la mise à feu et l'explosion, délai que l'on supposera gaussien. Lors de l'étude d'un certain type de dispositif d'allumage, les résultats des mesures effectuées sur un échantillon de taille 10, ont été, en millisecondes :

27, 28, 38, 32, 30, 28, 27, 30, 29, 31

Toutes les estimations par intervalle seront faites au niveau de confiance de 95%.

1. Supposons ici que  $\sigma = 3,2$ 
  - a. Déterminer l'intervalle de confiance de  $m$  et préciser son incertitude.
  - b. Quelle est la taille minimale de l'échantillon à considérer si on souhaite estimer  $m$  avec une incertitude n'excédant pas 1 milliseconde.

2. Supposons ici que  $\sigma$  est inconnu.

a.

i. Déterminer à nouveau l'intervalle de confiance de  $m$ .

ii. Comparer les deux intervalles de confiance obtenus et interpréter le résultat.

iii. Déterminer l'intervalle de confiance de  $\sigma$ .

b. Un dispositif d'allumage est conforme aux normes actuelles si son délai entre la mise à feu et l'explosion est inférieur à 32 millisecondes en moyenne.

A l'aide d'un test statistique, indiquer si le dispositif étudié est conforme ou pas ?

3. Dans cette question, on fournit les résultats des mesures effectuées sur un échantillon d'un autre dispositif d'allumage :

25 , 22 , 29 , 26 , 26 , 31 , 30 , 27

Le deuxième dispositif est-il plus performant ? (On pourra faire un test de comparaison de deux échantillons)

### Exercice 3. (7 points)

Dans un échantillon de 1000 personnes d'un certain pays, 40% sont des fumeurs. On observe, par ailleurs, que parmi les fumeurs, 175 personnes sont malades; et que parmi les non-fumeurs, 180 le sont.

- 1) Peut-on affirmer, que dans le pays considéré, la proportion de malades est la même chez les fumeurs que chez les non-fumeurs ?
- 2) l'Agence Nationale de Sécurité Sanitaire se demande si le pourcentage de fumeurs dans l'échantillon de taille 1000 considéré ci-dessus, est conforme à l'idée selon laquelle, la proportion de fumeurs à l'échelle du pays tout entier est supérieure à 35%.
  - a) Réaliser le test statistique devant aider l'Agence Nationale de Sécurité Sanitaire à répondre à sa préoccupation.
  - b) Quelle est alors la puissance de ce test ? Que traduit-elle ? (On choisira comme hypothèse alternative  $p_1 = 40\%$ )

# Examen de Statistique

HEI 3 - juin 2014

- *Durée : 2 heures*
- *Autorisés : calculatrice et recto-verso manuscrit*
- *Le barème n'est donné qu'à titre indicatif*

**Sauf mention contraire, on réalisera les tests au seuil de 5% en détaillant bien toutes les étapes : hypothèses, variable de décision et sa loi, région critique, valeur observée de la variable de décision et enfin conclusion**

## Exercice 1 (7 points).

1. On souhaite savoir s'il existe une dépendance entre la filière d'un étudiant en deuxième année de prépas et son sexe. Pour cela, on dispose des résultats suivants :

	PC	PSI	MP
M	70	96	86
F	80	13	23

Quelle est votre conclusion ? La réponse sera justifiée à l'aide d'un test du chi 2.

2. On souhaite maintenant étudier le lien entre la moyenne en Mathématiques d'un étudiant en deuxième année de prépas ( $X$ ) et sa moyenne générale ( $Y$ ). Pour cela, on dispose des moyennes suivantes calculées sur un échantillon de 54 étudiants ainsi que des résultats fournis par les options graphiques d'Excel et son utilitaire d'analyse (cf. Annexes).

	X	Y	ln X	ln Y
moyenne	10,44	11,14	2,33	2,41

- (a)
  - i. Pour le modèle linéaire, déterminer l'ordonnée à l'origine de la droite.
  - ii. Pour le modèle exponentiel, donner le coefficient de détermination.
  - iii. Pour le modèle puissance, compléter le résultat fourni par l'utilitaire d'analyse.
- (b) Quel modèle ajuste au mieux le nuage de points ?
- (c) En utilisant le meilleur des modèles, donner pour un étudiant ayant une moyenne de 11/20 en Mathématiques :
  - i. Une estimation ponctuelle de sa moyenne générale.
  - ii. Un intervalle de prévision de sa moyenne générale.

## Exercice 2 (7 points).

Les résultats d'une étude menée en 1967 sur 237 enfants, décrits par leur sexe et leur poids en livres (1 livre = 0.45 kg), sont résumés ci-dessous.

Distribution par classes suivant le poids des filles			Résumé des échantillons filles / garçons		
borne inf	borne sup	effectif		F	M
	74,5	10	moyenne	98,88	102,90
74,5	86,5	23	écart-type corrigé	18,62	19,09
86,5	98,5	21			
98,5	110,5	25			
110,5	122,5	23			
122,5		9			

1. Justifier, à l'aide d'un test du chi 2, que le poids des filles peut être modélisé par une loi normale.
2. En supposant que le poids des garçons peut, lui aussi, être modélisé par une loi normale, étudier si la différence entre le poids moyen des filles et celui des garçons est significative ou pas. On fera un test bilatéral pour comparer les moyennes.

Dans les deux exercices suivants, on utilisera les sorties Excel fournies que l'on complétera au besoin. Les différentes étapes des tests mis en oeuvre pour répondre aux questions ne sont pas attendues ici.

### Exercice 3 (2 points).

Les données suivantes correspondent à une expérience où 4 dentifrices (T1, T2, T3 et T4) ont été chacun testés sur 6 personnes afin que soit mesuré leur impact sur la blancheur des dents. Tous les patients utilisaient auparavant le même dentifrice.

T1	T2	T3	T4
16	18	19	20
17	20	27	23
17	20	28	24
19	21	29	25
21	22	32	26
24	23	34	29

Une ANOVA à un facteur fournit :

	A	B	C	D	E	F	G
1	Analyse de variance: un facteur						
2							
3	RAPPORT DÉTAILLÉ						
4	<i>Groupes</i>	<i>Nombre d'échantillons</i>	<i>Somme</i>	<i>Moyenne</i>	<i>Variance</i>		
5	T1	6	114	19	9,2		
6	T2	6	124	20,66666667	3,066666667		
7	T3	6	169	28,16666667	26,96666667		
8	T4	6	147	24,5	9,1		
9							
10							
11	ANALYSE DE VARIANCE						
12	<i>Source des variations</i>	<i>Somme des carrés</i>	<i>Degré de liberté</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Probabilité</i>	<i>Valeur critique pour F</i>
13	Entre Groupes						
14	A l'intérieur des groupes						
15							
16	Total						

Le dentifrice a-t-il un effet significatif sur la blancheur des dents ?

### Exercice 4 (4 points).

Pour étudier l'impact, sur le rendement, de la méthode de culture et du type de champ (mêmes sols, mais des expositions différentes), une expérience a été menée où trois méthodes de culture ont été testées sur trois types de champs.

1. Les rendements mesurés après la moisson sont regroupés dans le tableau suivant :

Méthode	Champ 1	Champ 2	Champ 3
1	20	39	34
2	35	30	58
3	62	82	69

Une ANOVA à deux facteurs sans répétition fournit :

	A	B	C	D	E	F	G
1	Analyse de variance: deux facteurs sans répétition d'expérience						
2							
3	<i>RAPPORT DÉTAILLÉ</i>	<i>Nombre d'échantillons</i>	<i>Somme</i>	<i>Moyenne</i>	<i>Variance</i>		
4		1	3	93	31	97	
5		2	3	123	41	223	
6		3	3	213	71	103	
7							
8	Champ 1		3	117	39	453	
9	Champ 2		3	151	50,33333333	772,3333333	
10	Champ 3		3	161	53,66666667	320,3333333	
11							
12							
13	ANALYSE DE VARIANCE						
14	<i>Source des variations</i>	<i>Somme des carrés</i>	<i>Degré de liberté</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Probabilité</i>	<i>Valeur critique pour F</i>
15	Lignes					0,025261596	
16	Colonnes						
17	Erreur	491,3333333					
18							
19	Total						

- (a) La méthode a-t-elle un effet significatif sur le rendement de la culture ?
- (b) Le type de champ a-t-il un effet significatif sur le rendement de la culture ?

2. En répétant l'expérience, les résultats suivants ont été obtenus :

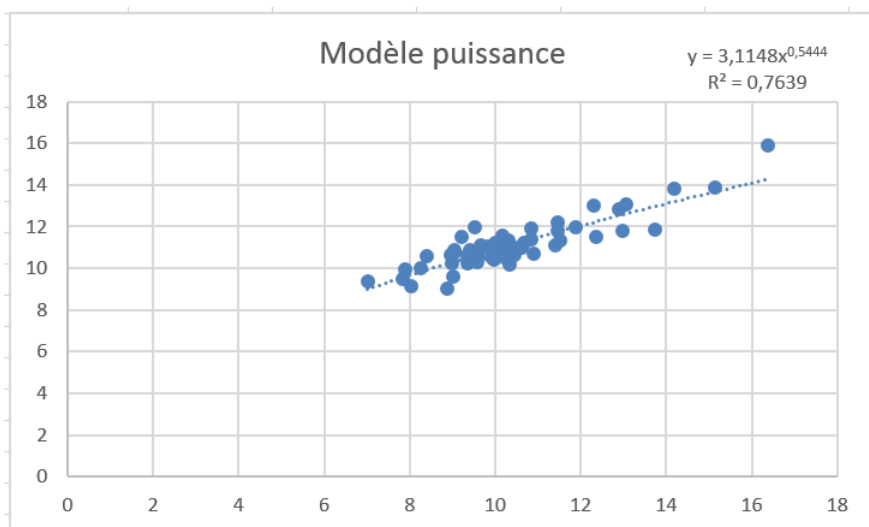
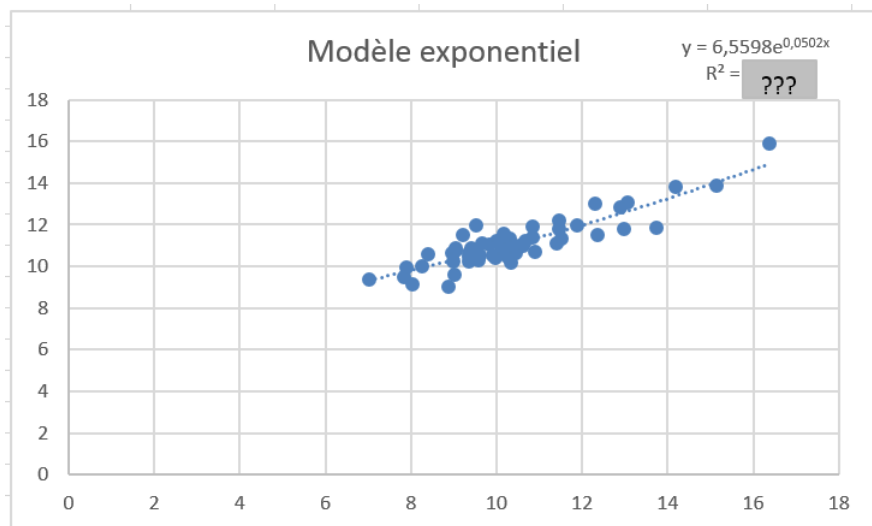
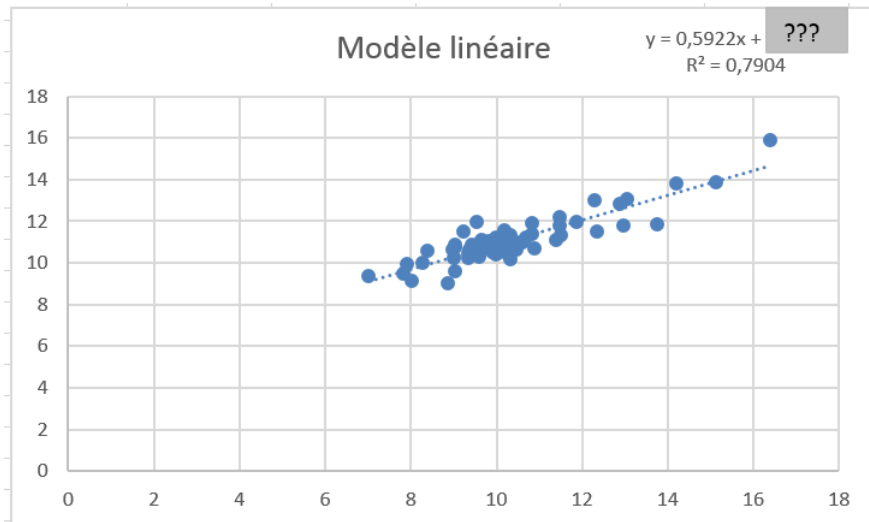
Méthode	Champ 1	Champ 2	Champ 3
1	20	39	34
	7	17	13
2	35	30	58
	52	28	73
3	62	82	69
	44	81	84

Une ANOVA à deux facteurs avec répétition fournit alors :

29	ANALYSE DE VARIANCE						
30	<i>Source des variations</i>	<i>Somme des carrés</i>	<i>Degré de liberté</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Probabilité</i>	<i>Valeur critique pour F</i>
31	Échantillon	7105,333333					
32	Colonnes	1027					
33	Interaction						
34	A l'intérieur du groupe	1081					
35							
36	Total	10684					

- (a) L'interaction entre la méthode et le type de champ a-t-elle un effet significatif sur le rendement de la culture ?
- (b) La méthode a-t-elle un effet significatif sur le rendement de la culture ?
- (c) Le type de champ a-t-il un effet significatif sur le rendement de la culture ?
- (d) Comparer les résultats obtenus aux questions 1.(b) et 2.(c).

## Annexe 1 : trois graphiques



## Annexe 2 : trois régressions linéaires

Régression linéaire entre X (variable explicative) et Y (variable expliquée)

	A	B	C
1	RAPPORT DÉTAILLÉ		
2			
3	<i>Statistiques de la régression</i>		
4	Coefficient de détermination multiple	0,889042772	
5	Coefficient de détermination R <sup>2</sup>	0,790397051	
6	Coefficient de détermination R <sup>2</sup>	0,786366225	
7	Erreur-type	0,564325256	
8	Observations	54	
9			
10	ANALYSE DE VARIANCE		
11		<i>Degré de liberté</i>	<i>Somme des carrés</i>
12	Régression	1	62,44680762
13	Résidus	52	16,56007571
14	Total	53	79,00688333
15			
16		<i>Coefficients</i>	<i>Erreur-type</i>
17	Constante		0,448125169
18	GR1 = X	0,592240519	0,042293389

Régression linéaire entre X (variable explicative) et ln Y (variable expliquée)

	A	B	C
1	RAPPORT DÉTAILLÉ		
2			
3	<i>Statistiques de la régression</i>		
4	Coefficient de détermination multiple	0,884335973	
5	Coefficient de détermination R <sup>2</sup>	0,782050113	
6	Coefficient de détermination R <sup>2</sup>	0,777858769	
7	Erreur-type	0,049075465	
8	Observations	54	
9			
10	ANALYSE DE VARIANCE		
11		<i>Degré de liberté</i>	<i>Somme des carrés</i>
12	Régression	1	0,449376256
13	Résidus	52	0,125236864
14	Total	53	0,57461312
15			
16		<i>Coefficients</i>	<i>Erreur-type</i>
17	Constante	1,8809562	0,038970347
18	GR1 = X	0,050239818	0,003677964

Régression linéaire entre ln X (variable explicative) et ln Y (variable expliquée)

	A	B	C
1	RAPPORT DÉTAILLÉ		
2			
3	<i>Statistiques de la régression</i>		
4	Coefficient de détermination multiple	0,874001963	
5	Coefficient de détermination R <sup>2</sup>	0,763879431	
6	Coefficient de détermination R <sup>2</sup>	0,759338651	
7	Erreur-type	0,051080249	
8	Observations	54	
9			
10	ANALYSE DE VARIANCE		
11		<i>Degré de liberté</i>	<i>Somme des carrés</i>
12	Régression	1	0,438935143
13	Résidus	52	0,135677977
14	Total	53	0,57461312
15			
16		<i>Coefficients</i>	<i>Erreur-type</i>
17	Constante	???	0,098104382
18	ln X	???	0,041973112